

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 08:36:32 ; Search time 87.3574 Seconds
(without alignments)
13131.604 Million cell updates/sec

Title: US-09-602-362E-15
Perfect score: 3403
Sequence: 1 ctggtgcgttaagatggt.....Gaataaaanaaaaaa 2030

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlp
-MODE=frame+n2p model -DEV=xlp
-O=/cgn2_1/USPTC spool_p/US0602362/runat_15072004_093625_21981/app_query.fasta_1.10325
-DB=A_Geneseq_25Jan04 -QMT=fascan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=humar40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09602362 @CGN 1.1.729 @runat_15072004_093625_21981 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELEXT=7

Database : A_Geneseq_25Jan04: *
1: Geneseqp1980s: *
2: Geneseqp1990s: *
3: Geneseqp2000s: *
4: Geneseqp2001s: *
5: Geneseqp2002s: *
6: Geneseqp2003as: *
7: Geneseqp2003bs: *
8: Geneseqp2004s: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|------------|--------------------|
| 1 | 3272 | 96.2 | 1341 | 4 AAB84702 | AAB84702 Amino aci |
| 2 | 3272 | 96.2 | 1341 | 5 ABJ05537 | ABJ05537 Breast ca |
| 3 | 3272 | 96.2 | 1341 | 6 ABJ37784 | ABJ37784 Human tum |
| 4 | 3272 | 96.2 | 1341 | 6 ABR47548 | ABR47548 Breast ca |
| 5 | 3272 | 96.2 | 1349 | 6 ABJ37788 | ABJ37788 Human tum |
| 6 | 3261 | 95.8 | 1002 | 4 AAU33351 | AAU33351 Human bre |
| 7 | 3261 | 95.8 | 1002 | 5 ABG78918 | ABG78918 Human bre |
| 8 | 3261 | 95.8 | 1002 | 6 ABJ37741 | ABJ37741 Human tum |
| 9 | 3261 | 95.8 | 1002 | 8 ADE44427 | ADE44427 Human bre |
| 10 | 3261 | 95.8 | 1095 | 4 AAU33357 | AAU33357 Human bre |

| | | | | | |
|----|--------|------|------|------------|--------------------|
| 11 | 3261 | 95.8 | 1095 | 5 ABG78924 | ABG78924 Human bre |
| 12 | 3261 | 95.8 | 1095 | 6 ABJ37747 | ABJ37747 Human tum |
| 13 | 3255 | 95.7 | 1013 | 6 ABJ37783 | ABJ37783 Human tum |
| 14 | 2599 | 76.4 | 512 | 4 AAB84701 | AAB84701 Amino aci |
| 15 | 2256.5 | 66.3 | 1239 | 6 ABJ37789 | ABJ37789 Human tum |
| 16 | 2184 | 64.2 | 445 | 4 AAB50249 | AAB50249 Human bre |
| 17 | 2184 | 64.2 | 445 | 4 AAG55987 | AAG55987 B726P spl |
| 18 | 2184 | 64.2 | 445 | 4 AAU33350 | AAU33350 Human bre |
| 19 | 2184 | 64.2 | 445 | 5 ABG78917 | ABG78917 Human bre |
| 20 | 2184 | 64.2 | 445 | 6 ABJ37740 | ABJ37740 Human tum |
| 21 | 2184 | 64.2 | 445 | 8 ADE44425 | ADE44425 Human bre |
| 22 | 2017 | 59.3 | 432 | 4 AAB50244 | AAB50244 Human bre |
| 23 | 2017 | 59.3 | 432 | 4 AAG55982 | AAG55982 B726P thi |
| 24 | 2017 | 59.3 | 432 | 4 AAU33345 | AAU33345 Human bre |
| 25 | 2017 | 59.3 | 432 | 5 AAO15779 | AAO15779 Human bre |
| 26 | 2017 | 59.3 | 432 | 5 ABG78912 | ABG78912 Human bre |
| 27 | 2017 | 59.3 | 432 | 6 ABJ37735 | ABJ37735 Human tum |
| 28 | 2017 | 59.3 | 432 | 8 ADE44133 | ADE44133 Human bre |
| 29 | 1785 | 52.5 | 1225 | 6 ABR47547 | ABR47547 Breast ca |
| 30 | 1597.5 | 46.9 | 650 | 4 AAB50263 | AAB50263 Human bre |
| 31 | 1597.5 | 46.9 | 650 | 4 AAG55983 | AAG55983 B726P spl |
| 32 | 1597.5 | 46.9 | 650 | 4 AAU33346 | AAU33346 Human bre |
| 33 | 1597.5 | 46.9 | 650 | 5 ABG78913 | ABG78913 Human bre |
| 34 | 1597.5 | 46.9 | 650 | 6 ABJ37736 | ABJ37736 Human tum |
| 35 | 1597.5 | 46.9 | 650 | 8 ADE44421 | ADE44421 Human bre |
| 36 | 1597.5 | 46.9 | 743 | 4 AAU33358 | AAU33358 Human bre |
| 37 | 1597.5 | 46.9 | 743 | 5 ABG78925 | ABG78925 Human bre |
| 38 | 1597.5 | 46.9 | 743 | 6 ABJ37748 | ABJ37748 Human tum |
| 39 | 1594.5 | 46.9 | 398 | 3 AAB07638 | AAB07638 Amino aci |
| 40 | 1591.5 | 46.8 | 661 | 6 ABJ37782 | ABJ37782 Human tum |
| 41 | 1546 | 45.4 | 1011 | 4 AAB84703 | AAB84703 Amino aci |
| 42 | 1514 | 44.5 | 335 | 4 AAB63900 | AAB63900 Human pro |
| 43 | 1404.5 | 41.3 | 410 | 4 AAU33359 | AAU33359 Human bre |
| 44 | 1404.5 | 41.3 | 410 | 5 ABG78926 | ABG78926 Human bre |
| 45 | 1404.5 | 41.3 | 410 | 6 ABJ37749 | ABJ37749 Human tum |

ALIGNMENTS

RESULT 1

AAB84702
ID AAB84702 standard; protein; 1341 AA.
XX
AC AAB84702;
XX
DT 17-SEP-2001 (first entry)
XX
DE Amino acid sequence of a human cancer associated antigen.
XX
KW Cancer associated antigen; INGI1; tumour suppressor; cancer; vaccine.
XX
OS Homo sapiens.
XX
PN WO200147959-A2.
XX
PD 05-JUL-2001.
XX
PF 29-NOV-2000; 2000WO-US042334.
XX
PR 30-NOV-1999; 99US-00451739.
PR 24-OCT-2000; 2000US-00602362.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
PA (SLOK) SLOAN KETTERING INST CANCER RES.
XX (CORR) CORNELL RES FOUND INC.

Jager D, Stockert E, Scanlan M, Knuth A, Old L, Gure A, Chen Y;

WPI; 2001-441706/47.

Isolated cancer associated nucleic acid molecule identified by SEREX
(serological identification of antigens by recombinant expression
cloning) technique, useful in nucleic acid based therapies to treat

PT cancer.
XX Claim 83; Page 53-57; 62pp; English.
XX The present sequence represents a human cancer associated antigen. The INGI
CC sequence was identified using probes derived from the INGI gene. The INGI
CC gene is a tumour suppressor candidate gene. The cancer associated antigen
CC polynucleotides and polypeptides are useful for screening for the
CC possible presence of a pathological condition in a subject such as
CC cancer. The cancer associated antigen polypeptides are useful for
CC producing vaccines
XX
SQ Sequence 1341 AA;

Alignment Scores:
Pred. No.: 2,04e-274 Length: 1341
Score: 3272.00 Matches: 656
Percent Similarity: 97.49% Conservative: 3
Best Local Similarity: 97.04% Mismatches: 13
Query Match: 96.15% Indels: 5
DB: 4 Gaps: 0

US-09-602-362E-15 (1-2030) x AAB84702 (1-1341)
QY 9 GTTAAAGATGGTCTTCTGAAGCTAACTCGGAATGAAGATTTCTATTCCACTAAGGCC 68
DB 666 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 685
QY 69 TTAGAATTGATGACATGCAAACTTTCAAGACAGAGCTCCCGAGAAGCCATCTGCCTTC 128
DB 686 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPhe 705
QY 129 GACCTCCCATGGAATGCAAAAGTCTGTTCCAAATTAAGACCTTGGAAATGAAGATGAA 188
DB 706 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 725
QY 189 CRAACATGAGACGACATGAGATCTCCATCAGATCCAAACAAAGAGCTATCAAGAA 248
DB 726 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 745
QY 249 AGTTCTTGGATTCGAGAGTCTCTGAGACTGTTTTCACAGAGATGTGTTTACCC 308
DB 746 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 765
QY 309 AAGCTTACACATCAAAAGAAATAGATAAATAATGAAATGAAATGAAAGCTCTCTGAT 368
DB 766 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 785
QY 369 AATGATGGTGTCTTGAAGGCTCCCTGCAGAAATGAAAGTTTCTATTCCAACTAAAGCCTTA 428
DB 786 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 805
QY 429 GAATGATGACATGCAAACTTCAAGACAGAGCTCCCGAGAGCCATCTCCCTCCGAG 488
DB 806 GluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu 825
QY 489 CTTGCCATTCGAAATGCAAAAGTCTGTTCCAAATTAAGACCTTGGAAATGAAGATGAACAA 548
DB 826 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 845
QY 549 ACATTGAGACGATCAGATGTTCCCTTCAGATCAAAACAAAGAGCTTGAAGAAAT 608
DB 846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLysValGluGluAsn 865
QY 609 TCTTGGATTCGAGACTCTCCGTGAGACTGTTTTCACAGAGGATGTGTGTGTTACCCCAAG 668
DB 866 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
QY 669 GCTACACATCAAAAGAAATGGAATAAATGAATGGAATAATGAAGATTCACATGACCTTA 728
DB 886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 905
QY 729 TCAAAATCTTGGATACAGTTCATTCTTGTGAAAGAGCAAGGAACTTCAAAAAGATCAC 788

DB 906 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 925
QY 789 TGTGAACAACGTACAGAAAAATGGAACAAATGAAAAAGATTTTGTCTACTGAAAAAG 848
DB 926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 945
QY 849 AAATCTGTCAAGCAAAAGAAATAAATCAGATGTAGAGAACCAAAAGTTAAATGGGAA 908
DB 946 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 965
QY 909 CAAGAGCTGTCAGTGTGAGATTGACTTTAAACCAAGAAAGAGAGAGAGAAATGCC 968
DB 966 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgAsnAla 985
QY 969 GATATATTAAATGAAAAATTAGGGAAGATATTAGGAAGATTCAGAGACGATAGGAAA 1028
DB 986 AspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGluGlnHisArgLys 1005
QY 1029 GAGTTAGAGTGAACAACAACACTTGAACAGCTCTCAGAAATCAAGATATAGAAATGGAAG 1088
DB 1006 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 1025
QY 1089 AGTGTGAAAGTAAATTTGAATCAGGTTTCTCACATCATGAAATGAAATATCTCTTA 1148
DB 1026 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 1045
QY 1149 CATGAAATTCATCTTGAAGAAAGAAATGCGATCTCAAACTGAAATAGCACACATG 1208
DB 1046 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysGluIleAlaThrLeu 1065
QY 1209 AAACACCAATACACAGGAAAGAAATTAATATCTTTGAGCACATTAAGATTTTAAAGAA 1268
DB 1066 LysHisGlnTyrGlnGlnLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 1085
QY 1269 AAGAACTGCTGAACCTTCAGATGACCTTAAACTCAAGAGAGAAATCAATTAAGAGGCA 1328
DB 1086 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluSerLeuThrLysArgAla 1105
QY 1329 TCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGAACACAATGTCTCACTTAAA 1388
DB 1106 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1125
QY 1389 TTGAAGCAAAACAAAGCAAAAGAAATCTAGAGGCGAAATTAATCAACACCTCTAGA 1448
DB 1126 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisProArg 1145
QY 1449 CTGCTTCTCTGTACAGACCATGATCAAAATGTGACATCAAGAAAGAAAGCTCAAGAACCT 1508
DB 1146 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1165
QY 1509 GCTTTCCACATTCGAGGAGATGCTTGTTCAAAGAAAAATGAATGTGTGATGTGAGTAGT 1568
DB 1166 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1185
QY 1569 ACCGATATATCAATCAGGTGCTCCATCAACACCTTCTGAAAGCTCAAGAGAAATCCAN 1628
DB 1186 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 1205
QY 1629 AAGCTTAAAAATTAATCTCAATTCGAGGAGATGCTTAAAGAGAAATACATGTGTTTC 1688
DB 1205 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 1225
QY 1689 AGGAACATGCAAAAGAGACCAAGCTGAAACAGAGTGTCAAAATCAAGAGAGCTGAACACA 1748
DB 1225 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHis 1245
QY 1749 TGTNTCAAAACGAAAGATNATGTGAACAAACACACTGANCAGAGAGTCTCTAGATC 1808
DB 1245 etTyrGlnAsnGluGlnAsnValAsnLysHisThrGluGlnGlnLysSerLeuAspG 1265
QY 1809 AGAAATATTTCACACTCAAGCAAAATATGTGGCTTCAACAGCAATAGTTCATCATCAC 1868

Db 1265 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnLeuValHisAlaH 1285
QY 1869 ATAANGAAGCTGCAACAAAGCAAGATACAAATTGATNTTCATNTTCTTGAGAGGAAA 1928
Db 1285 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1304
QY 1929 ATGCN-CATCATCTCTTAAAGAGAGAAAATGAGGAGATATTNATTACNATTAACCATTTA 1987
Db 1305 MetGlnHisLeuLeuLysGlnLysAsnGluGluIlePheAsnTrpAsnAsnHisLeu 1324
QY 1988 AAAAACCCGTATATTTCATATGGAAGAAAAAANAANA 2027
Db 1325 LysAsn-ArgIleTrpGlnTrpGluLysGluLysAlaGlu 1337
RESULT 2
AB05537
ID AB05537 standard; protein; 1341 AA.
AC ABJ05537;
XX
XX 14-NOV-2002 (first entry)
DE Breast cancer-associated protein 2.
XX
KW Breast cancer; breast cancer-associated gene sequence; drug development;
KW pharmacogenetics; biosensor development.
XX
OS Unidentified.
XX
XX WO200259377-A2.
PN
XX
XX 01-AUG-2002.
XX
XX 24-JAN-2002; 2002WO-US002242.
XX
XX 24-JAN-2001; 2001US-0263965P.
PR 02-FEB-2001; 2001US-0265928P.
PR 09-APR-2001; 2001US-00829472.
PR 09-APR-2001; 2001US-0282698P.
PR 04-MAY-2001; 2001US-0288590P.
PR 29-MAY-2001; 2001US-0294443P.
XX
XX (EOSB-) EOS BIOTECHNOLOGY INC.
XX
XX Mack DH, Gish KC, Afar D;
XX
XX WPI; 2002-583738/62.
DR N-PSDB; ABT07694.
XX
XX Detecting a breast cancer-associated transcript in a patient's cell,
PT useful for diagnosing breast cancer, comprises contacting a biological
PT sample with a polynucleotide that selectively hybridizes with breast
PT cancer nucleic acids.
XX
XX Disclosure; Page 348-349; 414pp; English.
PS
XX
XX The invention comprises a method of detecting a breast cancer-associated
CC transcript in a cell from a patient. The method of the invention involves
CC contacting a biological sample from the patient with a nucleotide that
CC hybridizes to one of the 69 breast cancer-associated gene sequences shown
CC in the specification. The method of the invention is useful in the
CC diagnosis or prognosis of breast cancer, and for detecting genes that are
CC up or down-regulated in breast cancer cells. Genes identified by the
CC method of the invention can be used in diagnostic purposes and also as
CC targets for screening for therapeutic compounds that modulate breast
CC cancer (e.g. hormones or antibodies). Identification of genes that are
CC over or under expressed in breast cancer can additionally provide high-
CC resolution, high-sensitivity datasets which can be used in the areas of
CC diagnostics, therapeutics, drug development, pharmacogenetics, protein
CC structure and biosensor development. Amino acid sequences ABJ0536 -
CC ABJ05604 represent the proteins encoded by the 69 breast cancer-
CC associated genes of the invention
XX

SQ Sequence 1341 AA;
Alignment Scores:
Pred. No.: 2,04e-274 Length: 1341
Score: 3272.00 Matches: 656
Percent Similarity: 97.49% Conservative: 3
Best Local Similarity: 97.04% Mismatches: 13
Query Match: 96.15% Indels: 5
DB: Gaps: 0
US-09-602-362E-15 (1-2030) x ABJ05537 (1-1341)
QY 9 GTTAAAGATGCTCTTCTGAAGGCTAACTGCGGAATGAAAGTTTCTATTCCAACTAAAGCC 68
Db 666 VallysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 685
QY 69 TTAGAATTGATGACATGCAAACTTTCAAGACAGAGCCTCCCGAGAGCCATCTGCCTTC 128
Db 686 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPhe 705
QY 129 GAGCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCCTTGAATTGAAGATGAA 188
Db 706 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 725
QY 189 CAAACATTGAGACAGATGAGATACCTCCCATCAGAAATCCAAACAAAGACATATGAAGAA 248
Db 726 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTrpGluGlu 745
QY 249 AGTCTCTGGATTCTGAGATCTCTGAGACTGTTTCACAGAGGATGTGTGTTTACCC 308
Db 746 SerSerTrpAspSerGluSerLysGluThrValSerGlnLysAspValCysLeuPro 765
QY 309 AAGCTACACATCAAAAGAAATAGATAAAATAATGAAATGAAATAGAAAGTCTCTGAT 368
Db 766 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 785
QY 369 AATGATGGTTTCTGAAGGCTCCTCGAGATGAAATGAAATGAAATGAAATGAAATGAAAT 428
Db 786 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 805
QY 429 GAATGATGACATGCAAACTTTCAAAGACAGCCTCCCGAGAGCCATCTGCCTTCGAG 488
Db 806 GluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu 825
QY 489 CTTGCCATTCAAATGCAAAAGTCTGTTCCAAATGAAAGCTTGAATGAAATGAAATGAAAT 548
Db 826 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 845
QY 549 ACATTGAGACAGATCAGATGTTCCCTTCAGATCAAAACAAAGAGGTTGAAGAAAT 608
Db 846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsn 865
QY 609 TCTTGGATTCGAGAGTCTCCGTCGAGACTGTTTCACAGAGGATGTGTGTGTACCCAG 668
Db 866 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
QY 669 GCTACACATCAAAAGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAAT 728
Db 886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 905
QY 729 TCAAAATCTTGGATACAGTTTCATTTCTGTAAGAGCAAGGCACTTCAAAAAGATCAC 788
Db 906 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 925
QY 789 TGTGAACAGCTACAGAAAATGGAACAAATGGAACAAATGGAACAAATGGAACAAATGGAAC 848
Db 926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 945
QY 849 AAATCTGTCAGAACCAAGAAATATAATACACTTAGAGAACCAAGAAATGTAATGGAA 908
Db 946 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 965
QY 909 CAAGAGCTCTGCAGTGTGAGATTGATTTAAACCAAGAGAGAGAGAGAGAAATGCC 968

QY 189 CAAACATTGAGACGATGAGATCTCCCATCAGAAATCAACAAAAGGACTATGAGAA 248
 Db 726 GlnThrLeuArgAlaAspGluLeuProSerGluSerLysGlnLysAspTyrGluGlu 745
 QY 249 AGTTCTTGGGATCTGAGAGTCTCTGAGACTGTTTCACAGAGGATGTTGTTACCC 308
 Db 746 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 765
 QY 309 AAGGCTACACATCAAAAAGAAATAGATAATAAATGAAAATAGAGAGTCTCTCTGAT 368
 Db 766 LysAlaThrHisGlnLysGluLeuAspLysIleAsnGlyLysLeuGluGluSerProAsp 785
 QY 369 AATGATGGTTTCTGAGAGCTCCCTCGCAGATGAAAGTTTCTATCCAACTAAAGCCCTTA 428
 Db 786 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 805
 QY 429 GAATGTGAGCATCAAACTTTCAAAGCAGAGCTCCCGAAGCCATCTGCTTCGAG 488
 Db 806 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 825
 QY 489 CTGCAATTGAAATGCAAAAGTCTCTTCCAAATAAAGCTTTGGAATGGAAGATGAACAA 548
 Db 826 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 845
 QY 549 ACATTGAGACGATCAGATGTTCTCCCTTCAGATCAAAACAAAGAGGTTGAGAAAT 608
 Db 846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsn 865
 QY 609 TCTTGGGATCTGAGAGTCTCCGTCGAGACTGTTTCACAGAGGATGTTGTTGACCCAAAG 668
 Db 866 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
 QY 669 GCTACATCAAAAAGAAATGATGATAAATAGTGAATAATAGAGATTCACATGACCTA 728
 Db 886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerTrpSerLeu 905
 QY 729 TCAAAAATCTTGGATACAGTTCATCTTGTGAAGACCAAGGAACTTCAAAAAGATCAC 788
 Db 906 SerLysIleLeuAspThrValHisSerCysGluArgAlaGluLeuGlnLysAspHis 925
 QY 789 TGTGAACACGTACAGAAAATGAACAATGAACAAAGATTTTGTGTACTGAAAAG 848
 Db 926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys 945
 QY 849 AAACGTGTGAGCAAGCAAAAGAAATAAATCAAGTTAGAGAACCAAAAAGTTAAATGGGAA 908
 Db 946 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 965
 QY 909 CAAGAGCTCTGAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAGAAATGCC 968
 Db 966 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 985
 QY 969 GATATATTAAATCAAAAATTTAGGGAAGAAATAGGAAGAAATCGAAGACCATAGGAA 1028
 Db 986 AspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGluGlnHisArgLys 1005
 QY 1029 GAGTTAGAGTGAACCAACACTTCAACAGGCTCTCAGAAATCAAGATATAGAAATGAAG 1088
 Db 1006 GluLeuGluValLysGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 1025
 QY 1089 AGTGTAGAAAGTAATTTGATGATCAGTTTCTCAGCTCATGAAATGAAATATATCTCTTA 1148
 Db 1026 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTrpLeuLeu 1045
 QY 1149 CATGAAATTTGATGTGAAAAGGAAATTTGCATGCTTAAACTGGAATAGCCACACTG 1208
 Db 1046 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 1065
 QY 1209 AAACACCAATACCAAGAAAAGAAAATAAATCTTTGAGGACATTAAGATTTTAAAGAA 1268
 Db 1066 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 1085

QY 1269 AAGAATGCTGAACCTTCAGATGACCCCTAAAACTGAAAGAGAAATCATTAATAAAGGCA 1328
 Db 1086 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluSerLeuThrLysArgAla 1105
 QY 1329 TCTCAATATAGTGGGAGCTTAAAGTTCATGATCTGAGACCAACAATGCTCCTCTTCAA 1388
 Db 1106 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1125
 QY 1389 TTGAAGGAAAACAAGACAAAGAAATACTAGAGCGAGAAATTTGAATCACACCATCTAGA 1448
 Db 1126 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 1145
 QY 1449 CTGGCTTCTCTCTACAGACCATGATCAAAATGTGACATCAAGAAAAGTCAAGACCT 1508
 Db 1146 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1165
 QY 1509 GCTTTCCACATTCAGAGAGATGCTTTGTCAAAAGAAAATGAATGTGTGAGTAGT 1568
 Db 1166 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1185
 QY 1569 ACCGATATATAACAATGAGGTGCTCCATCAACCATCTTCTGAAGCTCAAGGAAATCCAN 1628
 Db 1186 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 1205
 QY 1629 AAGCTAAAATTAATCTCAATTCAGAGAGATGCTTAAAGAGAAAATACATTGGTTTC 1688
 Db 1205 SerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 1225
 QY 1689 AGAAACATGCAAAAGACCAACAGTGAACACACAGTGTCAAATGAAGGAGCTGTAGATC 1748
 Db 1225 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 1245
 QY 1749 TGTTTCAAAACGACCAAGATNATGTGAACAAACACACACTGANCAGCAGGAGTCTCTGATC 1808
 Db 1245 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1265
 QY 1809 AGAAATTAATTCAACTACAAAGCAAAAATATGTGGCTTCAACAGCAATTAGTTTCATGCCAC 1868
 Db 1265 LysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnLeuValHisAlaH 1285
 QY 1869 ATAANGAAGCTGACAAACAAACCAAGATTAACAATTGATTTTCAATTTCTTGAGAGGAA 1928
 Db 1285 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1304
 QY 1929 ATGCN-CATCATCTCTTAAAGAGAAAATGAGGAGATATTNNATTACNATAACCATTTA 1987
 Db 1305 MetGlnHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1324
 QY 1988 AAAAACCCGTATATTCATATTCGAAAAAANAANAANA 2027
 Db 1325 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1337
 RESULT 4
 ABR47548
 ID ABR47548 standard; protein; 1341 AA.
 XX AC ABR47548;
 XX AC
 DT 12-JUN-2003 (first entry)
 XX XX
 DE Breast cancer associated protein sequence SEQ ID NO:334.
 XX KW Human; breast cancer; cytostatic; gene therapy.
 XX OS Homo sapiens.
 XX PN WO2003004989-A2.
 XX XX
 PD 16-JAN-2003.
 XX XX
 PF 21-JUN-2002; 2002WO-US019669.
 XX XX
 PR 21-JUN-2001; 2001US-0299887P.

PR 27-JUN-2001; 2001US-0301572P.
 PR 18-JUL-2001; 2001US-0306501P.
 PR 25-SEP-2001; 2001US-0325002P.
 PR 05-MAR-2002; 2002US-0362585P.
 PR 14-MAY-2002; 2002US-0380391P.
 XX (MILL-) MILLENIUM PHARM INC.
 XX
 XX Lillie J, Gannavarapu M, Glatt K, Hoerish S, Kamatkar S;
 PI Mertens M, Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE;
 PI Bast RC, Hortobagyi GN, Pusztai L, Meric F, Sahin A, Mills GB;
 XX
 XX WPI; 2003-210381/20.
 DR N-PSDB; ACC50246.
 XX
 XX Breast cancer diagnosis or treatment by comparing the level of expression
 PT of a marker in a patient sample with that in the control non-breast
 PT cancer sample.
 XX
 PS Claim 1; SEQ ID NO 334; 128pp; English.
 XX
 XX The present invention describes a method for assessing whether a patient
 CC is afflicted with breast cancer. The method comprises comparing the level
 CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
 CC ABR47386 to ABR47632) in a patient sample and the normal level of
 CC expression of the marker in a control non-breast cancer sample, where a
 CC significant increase in the level of expression of the marker in the
 CC patient sample and the normal level is an indication that the patient is
 CC afflicted with breast cancer. The breast cancer associated sequences from
 CC the present invention have cytostatic activities and can be used in gene
 CC therapy. The method is useful for diagnosing and treating breast cancer.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1341 AA;
 Alignment Scores:
 Pred. No.: 2,04e-274 Length: 1341
 Score: 3272.00 Matches: 656
 Percent Similarity: 97.49% Conservative: 3
 Best Local Similarity: 97.04% Mismatches: 13
 Query Match: 96.15% Indels: 5
 DB: 6 Gaps: 0
 US-09-602-362E-15 (1-2030) x ABR47548 (1-1341)
 QY 9 GTTAAAGATGGTCTCTCGAAGCTAACTCCGGAATGAAGATTCTATTCCAACTAAAGCC 68
 DB 666 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 685
 QY 69 TTAGATTGATGACATGCAAACTTTCAAGCAGAGCCCTCCGAGAAGCCATCTCCCTTC 128
 DB 686 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 705
 QY 129 GAGCCTGCCATTGAAATGCAAAAGTCTGTTCAAAATAAAGCCTTGGAATTGAAGATGAA 188
 DB 706 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLysAsnGlu 725
 QY 189 CAACACTTCAGACATGATGATCTCCCATCAGATCCAAACAAAGGACTATCAAGAA 248
 DB 726 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspThrGluGlu 745
 QY 249 AGTTCTTGGGATCTCAGAGTCTCTGTGAGACTGTTTCACAGAAGGATGTGTGTTACCC 308
 DB 746 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 765
 QY 309 AAGCTTACATCAACAAAGAAATAGNTAAATAATGGAATTAAGAGTCTCTCTGAT 368
 DB 766 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 785
 QY 369 AATGATGGTCTTCTGAGGCTCCCTGCAGATGAAAGTTTCTATTCCAACTAAAGCCCTTA 428

786 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 805
 QY 429 GAATTGATGACATGCAAACTTTCAAAGCAGAGCCCTCCCGAGAAGCCATCTGCTTCGAG 488
 DB 806 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 825
 QY 489 CTGCGCATTCGAAATGCAAAAGTCTGTTCAAAATAAAGCCTTGGAATTGAAGATGAACAA 548
 DB 826 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLysAsnGluGln 845
 QY 549 ACATTGAGACACATGATGTTCCCTTCAGATCAAAACAAAGAGAGTGTGAAGAAAT 608
 DB 846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsn 865
 QY 609 TCTTGGATTCTCAGAGTCTCCGTGAGACTGTTTCACAGAAGGATGTGTGTACCCAAG 868
 DB 866 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
 QY 669 GCTACACATCAAAAGAAATGGAATAAAGTGAAGAAATAGAGATTCACATCACTACCTA 728
 DB 886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 905
 QY 729 TCAAAAATCTGGATACAGTTCATTCTGTGAAAGAGCAAGGAACTTCAAAAAGATCAC 788
 DB 906 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGluLysAspHis 925
 QY 789 TGTGAACAACGTACAGGAAATGGAACAATAAGAAAGAGTGTGTGTACTGAAAAG 848
 DB 926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 945
 QY 849 AAATCTGCAGAAGCAAAAGAAATAAATCACAGTTAGAGAACCAAAAGCTTAAATGGGAA 908
 DB 946 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 965
 QY 909 CAAGAGCTTCGAGTGTGAGATTGATTTAAACCAAGAGAGAGAGAGAGAAATGCC 968
 DB 966 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 985
 QY 969 GATATATTAATCAAAATTTAGGGAAGATTAAGGAAGATCAAGAGAGATAGCAAA 1028
 DB 986 AspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGluGlnHisArgLys 1005
 QY 1029 GAGTTAGAACTGAAACCAACTTGAACAGGCTCTCAGAAATACAGATATAGAAATGAAG 1088
 DB 1006 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 1025
 QY 1089 AGTGTAGAAATTAATTTGATCAGGTTCTTCACACTCATCAAAATGAAATATATCTCTTA 1148
 DB 1026 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnThrLeuLeu 1045
 QY 1149 CATGAAATTCATGTTGAAAAAGAAATTCGCATCTCAAACTGAAATAGCAACATG 1208
 DB 1046 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 1065
 QY 1209 AAACACCAATACAGGAAAGAAATAATATCTTTGAGACATTAAGATTTTAAAGAA 1268
 DB 1066 LysHisGlnThrGlnGluLysGluAsnLysThrPheGluAspIleLysIleLeuLysGlu 1085
 QY 1269 AAGAATCTCAACTTCAGATGACCTCAAACTCAAAAGAGAGAAATCAATACTAAAGGGCA 1328
 DB 1086 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysAspAla 1105
 QY 1329 TCTCAATATAGTGGGAGCTTAAAGTCTTGATAGCTGAGACACATGCTCACTTCTAA 1388
 DB 1106 SerGlnThrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1125
 QY 1389 TTGAAGGAAAAACAAGCAAAAGAAATACTAGAGCCAGAAATTAATCAACACCATCTCAGA 1448
 DB 1126 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 1145
 QY 1449 CTGCTTCTCTGTACAGACCATGATCAAAATTTGTGACATCAAGAAAGTCAAGAACCT 1508
 DB 1146 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1165

QY 1509 GCTTCCACATTCGAGGAGCTGTTGTTGCAAGAGAAAAATGAATGTTGATGTAGTAGT 1569
 Db 1166 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1185
 QY 1569 ACCGATATATACATGAGTCTCCATCAACCACTTCTGAGCTCAAGAGAAATCCAN 1628
 Db 1186 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLy 1205
 QY 1629 AAGCCTAAAAATTAATCTCAATATGTCAGGAGATGCTCTAAGAGAAAAATACATGTTTC 1688
 Db 1205 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 1225
 QY 1689 AGGACATCCACAAAGAGACCACTGTAACACACAGCTGTCAAATGAGGAGCTGAACACA 1748
 Db 1225 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 1245
 QY 1749 TGTNTCAAAACCAAGATNATGCAACAAACACACTCANAGCAGGAGTCTCTAGATC 1808
 Db 1245 erTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGlnSerLeuAspG 1265
 QY 1809 AGAATATTTCACTACAAAGCAAAATATATGTCGCTCAACAGCAATAGTTCATGCAC 1868
 Db 1265 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnValHisAlaH 1285
 QY 1869 ATAANGAAAGCTGACAAACAAAGCAAGATAACAATGATNTTTCATTNTCTTGAGAGAAA 1928
 Db 1285 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1304
 QY 1929 ATGCN-CATCATCTCTTCTAAAGAGAAAAATAGGAGATATTNATTACNATAACCATTTA 1987
 Db 1305 MetGlnHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1324
 QY 1988 AAAAAACCGTATATTCATATGGAAGAAAAAANAAAAA 2027
 Db 1325 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1337
 RESULT 5
 ID ABJ37788
 AC ABJ37788 standard; protein; 1349 AA.
 XX AC ABJ37788;
 XX DT 15-MAY-2003 (first entry)
 XX DE Human tumour-related protein - SEQ ID No 573.
 XX KW Human; vaccine; gene therapy; T cell stimulation; T cell expansion;
 XX KW tumour; breast cancer; cancer; immune response stimulation.
 XX OS Homo sapiens.
 XX PN WO200283956-A1.
 XX PD 24-OCT-2002.
 XX PF 15-APR-2002; 2002WO-US012378.
 XX PR 13-APR-2001; 2001US-00834759.
 XX PR 07-DEC-2001; 2001US-00007805.
 XX PR 13-FEB-2002; 2002US-00076622.
 XX XX (CORI-) CORIXA CORP.
 XX XX Houghton RL, Sleath PR, Persing DH, Jiang Y, Dillon DC;
 PI Mitcham JL, Xu J, Harlocker SL, Hepler WT, Henderson RA, Fanger GR;
 PI Vedvick TS, McNeill PD, Durham M;
 XX WPI; 2003-103376/09.
 DR New polypeptide and polynucleotide useful for stimulating and/or
 XX expanding T cells specific for a tumor protein and treating breast
 PT cancer.

XX Example 12; Page 353-357; 375pp; English.
 XX The invention comprises a method of stimulating and/or expanding T cells
 CC specific for a tumour protein. The invention further comprises human
 CC nucleic acids and proteins that are associated with tumours (e.g. breast
 CC cancer). The method and sequences of the invention are useful for
 CC stimulating and/or expanding T cells specific for a tumour protein,
 CC detecting the presence of cancer, stimulating an immune response in a
 CC patient and treating breast cancer. The present amino acid sequence
 CC represents a human tumour-related protein
 XX Sequence 1349 AA;
 SQ
 Alignment Scores:
 Pred. No.: 2,04e-274 Length: 1349
 Score: 3272.00 Matches: 656
 Percent Similarity: 97.49% Conservative: 3
 Best Local Similarity: 97.04% Mismatches: 13
 Query Match: 96.15% Indels: 5
 DB: Gaps: 0
 US-09-602-362E-15 (1-2030) x ABJ37788 (1-1349)
 QY 9 GTTAAAGATGCTTCTCTGAAGGCTAACTGCGGAATGAAAGTTCCTTATTCCAACTAAAGCC 68
 Db 674 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 693
 QY 69 TTGAATGTATGACATGCAAACTTTTCAAGCAGAGCCTCCCGAGAGCCATCTGCTTC 128
 Db 694 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 713
 QY 129 GAGCCTGCCATTCGAATGCAAAAGTCTGTTCCAAATAAAGCCCTGGAAATGAAGATGAA 188
 Db 714 GluProAlaLeuGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 733
 QY 189 CAAACATGTAGCAGATGAGATCTCCATCAGAATCCAAACAAAGGACTATGAAGAA 248
 Db 734 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 753
 QY 249 AGTCTCTGGATTCGAGATCTCTGAGACTGTTTCAAGAGAGATGCTGCTTACCC 308
 Db 754 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 773
 QY 309 AAGCTACACATCAAAAGAAATAGATAAAATAAATGGAATAATAGAAAGTCTCTGAT 368
 Db 774 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluSerProAsp 793
 QY 369 AATGATGTTTCTGAAGGCTCCTCGAGATGAAAGTTCCTTATTCCAACTAAAGCTTA 428
 Db 794 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 813
 QY 429 GAATGTATGACATGCAAACTTTCAAGCAGAGCCTCCCGAGAGCCATCTGCTTCGAG 488
 Db 814 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 833
 QY 489 CCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCCTGGAATGGAAGATGAACAA 548
 Db 834 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 853
 QY 549 ACATTGAGACGACATCAGATGTTCCCTTCAGATCAAAACAAAGAGGTTGAGAGAAAT 608
 Db 854 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsn 873
 QY 609 TCTTGGATTCGAGATCTCCCTGAGACTGTTTCAAGAGAGATGTTGTGTACCCAAAG 668
 Db 874 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 893
 QY 669 GCTACACATCAAAAGAAATGGAATAAAGTGGAAATAGAAATTCAGAAATTCAGCTA 728
 Db 894 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 913
 QY 729 TCAAAATCTTGGATACAGTTCACTTCTGTGAAAGAGCAAGGGAAGTTCAGAAAGATCAC 788

Db 914 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 933
QY 789 TGTGAACAACTGACAGGAAATGGAACAAATGAAAGAGAGTGTGTACTGAAAG 848
Db 934 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 953
QY 849 AACTGTGACAGCAAGCAAAATATAATCACAGTTAGAGAACCAAAAGTAAATGGGAA 908
Db 954 LysLeuSerGluAlaLysGluLeuLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 973
QY 909 CAAGAGCTCTGAGTGTGAGTGAATTTAAACCAAGAAAGAGAGAGAGAAATGCC 968
Db 974 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluArgLysArgArgAsnAla 993
QY 969 GATATATTAAATGAAGAAATAGGAAGAAATAGGAAGAAATAGGAAGCAATAGGAAA 1028
Db 994 AspIleLeuAsnGlnLysIleArgGluLeuGluArgLysGlnGluGlnHisArgLys 1013
QY 1029 GAGTTAGAAAGTGAACCAAACTTGAAACAGGCTCTCAGAAATACAGATATAGAAATGAAG 1088
Db 1014 GluLeuGluValLysGlnGlnLeuGlnAlaLeuArgIleGlnAspIleGluLeuLys 1033
QY 1089 AGTGTAGAAGTAATTTGATCAGCTTCTCAGCTCATGAAATCAAAATATCTCTTA 1148
Db 1034 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 1053
QY 1149 CATGAAATTTGCAATGTTGAAAAAGGAAATGTCATGCTAAAACTGGAATAGCCACACTG 1208
Db 1054 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluLeuAlaThrLeu 1073
QY 1209 AAACACAAATACAGGAAAGGAAATATAATCTTTGAGGACATTAAGATTTAAAGAA 1268
Db 1074 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 1093
QY 1269 AGAAATGCTGAATCTCAGATGACCTTAAACCTGAAAGAGCAATCAATTAATAAGGGCA 1328
Db 1094 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 1113
QY 1329 TCTCAATATAGTGGCAGCTTAAAGTCTGATAGCTGAGAACACAAATGCTCTCTTAA 1388
Db 1114 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1133
QY 1389 TTGAAGCAAAACACAGCAAGAAATAGTACAGGAGCAATGAATCACCACATCTCTAGA 1448
Db 1134 LeuLysGlnLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 1153
QY 1449 CTGGCTTCTGCTGACAGACCATGATCAATTTGACATCAAGAAAAGTCAAGAACT 1508
Db 1154 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1173
QY 1509 GCTTTCCACATTCAGAGAGTGTCTTTCGAAGGAAATGAATCTTCATGTGAGTAGT 1568
Db 1174 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1193
QY 1569 ACCGATATATAAATCAGAGTGTCTTCAATCAACACCTTCTGAGCTCAAGGAAATCCAN 1628
Db 1194 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 1213
QY 1629 AAGCTTAAATTAATCTCAATATGACAGAGATGCTCTAAGAGAAATACATTTGTTTC 1688
Db 1213 SerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSer 1233
QY 1689 AGGAAATGACAAAGAGACCAACAGTGAACACAGTGTCAATGAAGGAGTGAACACA 1748
Db 1233 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 1253
QY 1749 TGTNTCAAACCAAGATATATGTGAACAAACACACTGANCAGCAGGAGTCTTAGATC 1808
Db 1253 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1273
QY 1809 AGAAATTTTCAACTACAAGCAAAATATGTGGCTTCAACAGCAATTTAGTTTCATGCAC 1868

Db 1273 InLysLeuPheGlnLeuGlnSerLysAsnMetTyrLeuGlnGlnGlnLeuValHisAlaH 1293
QY 1869 ATRANGAAAGCTCACAAACAAAGCAAGATAACAAATTCATTTTCATTTGAGAGGAAA 1928
Db 1293 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1312
QY 1929 ATGCN-CATCATCTTCTAAAGAGAAATAGAGAGATATTNNATTACNATACCATTTA 1987
Db 1313 MetGlnHisHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1332
QY 1988 AAAAAACCCGATATTTCAATATGCAAAAAAANAANAANA 2027
Db 1333 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1345
RESULT 6
AAU33351
ID AAU33351 standard; protein; 1002 AA.
XX AC AAU33351;
XX 18-DEC-2001 (first entry)
XX Human breast cancer protein B726P from alternatively spliced cDNA.
XX Human; ss; breast cancer protein; tumour; cancer; cytostatic;
XX Gene therapy; immunogen.
XX Homo sapiens.
XX WO200179286-A2.
XX 25-OCT-2001.
XX 12-APR-2001; 2001WO-US012164.
XX 17-APR-2000; 2000US-00551621.
XX 08-JUN-2000; 2000US-00590751.
XX 22-JUN-2000; 2000US-00604287.
XX 20-JUL-2000; 2000US-00620405.
XX (CORI-) CORIXA CORP.
XX Jiang Y, Dillon DC, Mitcham JL, Xu J, Harlocker SL, Hepler WT;
XX WPI; 2001-611721/70.
XX N-PSDB; AAS47411.
XX Breast Tumor Proteins and nucleic acids useful for the prevention,
XX diagnosis and treatment of breast cancer.
XX Claim 3; Page 281-283; 297pp; English.
XX The invention relates to isolated breast tumour proteins and nucleic
XX acids that encode them, including immunogenic fragments of the proteins.
XX Also included are expression vectors expressing the proteins, transformed
XX cells and antibodies raised against the proteins or an antigen presenting
XX cell expressing the protein. The proteins and nucleic acids may be used
XX in the prevention, diagnosis and treatment of diseases associated with
XX inappropriate breast tumour protein expression, i.e. breast tumours and
XX breast cancer e.g. by gene therapy. The nucleic acids and their
XX complements may also be used as DNA probes in diagnostic assays to detect
XX and quantitate the presence of similar nucleic acids in samples, and
XX therefore which patients may be in need of restorative therapy. The
XX proteins, nucleic acids and antibodies may be used in assays to identify
XX modulators (e.g. antagonists) of breast tumour protein expression and
XX activity. The antibodies and antagonists may also be used to down
XX regulate expression and activity. The antibodies may also be used as
XX diagnostic agents for detecting the presence of the proteins in samples
XX (e.g. by enzyme linked immunosorbent assay (ELISA)) and in other immuno-
XX purification diagnostic techniques. The present sequence is a breast
XX tumour protein encoded by a cDNA from a breast tumour cDNA library
XX isolated by subtractive hybridisation against a normal breast CDNA
XX library

| | | |
|--|-----------|--|
| XX | SQ | Sequence 1002 AA; |
| Alignment Scores: | | |
| Pred. No.: | 1,66e-273 | Length: 1002 |
| Score: | 3261.00 | Matches: 654 |
| Percent Similarity: | 97.19% | Conservative: 3 |
| Best Local Similarity: | 96.75% | Mismatches: 15 |
| Query Match: | 95.83% | Indels: 5 |
| DB: | 4 | Gaps: 0 |
| US-09-602-362E-15 (1-2030) x AAU33351 (1-1002) | | |
| QY | 9 | GTATAAGATGCTTCTCGAAGCTTAACCTCGGAATCAAGTTTCTATTCCAACTAAAGCC 68 |
| DB | 327 | VAllyAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 346 |
| QY | 69 | TTAGAATTGATGCACATGCAAACTTTCAAAGCAGAGCCTCCCGAGAAGCCCATCTGCCCTTC 128 |
| DB | 347 | LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPhe 366 |
| QY | 129 | GAGCCTGCCATTGAATGCAAAAGCTGTCTCCAATAAAGCCTTGGAATTCAGAAGATGAA 188 |
| DB | 367 | GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 386 |
| QY | 189 | CAAACTTGCAGCACATGAGATACTCCCATCAGAACTCAAAACAAAAGGACTATGAGAA 248 |
| DB | 387 | GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrgluGlu 406 |
| QY | 249 | AGTCTTGGATTCTCGAGAGTCTGTGTGAGACTGTTTACAGAAGATGTGTGTTACC 308 |
| DB | 407 | SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 426 |
| QY | 309 | AAGCCTACACATCAAAAGAAATAGATAAATAAATGGAATTAGAAGAGTCTCTGAT 368 |
| DB | 427 | LysAla***HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProasp 446 |
| QY | 369 | AATGATGGTTTTCTGAAGGCTCCCTGCAGAATGAAAGTTTCTATTCCAACTAAAGCCCTTA 428 |
| DB | 447 | AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 466 |
| QY | 429 | GAATTGATGCATGCAAACTTTCAAAGCAGAGCCTCCGAGAAGCCATCTGCCTTCGAG 488 |
| DB | 467 | GluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu 486 |
| QY | 489 | CCTGCCATTGAAATGCAAAAGTCTGTTCAAAATAAAGCCTTGCgaattgaaGaAtGAACA 548 |
| DB | 487 | ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 506 |
| QY | 549 | ACATTGAGACGATCAGATGTTCCCTTCAGAAATCAAAACAAAAGAGGTTGAGAAAAAT 608 |
| DB | 507 | ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys**ValGluGluAsn 526 |
| QY | 609 | TCTTGGATTCTCAGAGTCTCCGTGAGACTGTTTCACAGAAGAGTGTGTGTACCCCAAG 668 |
| DB | 527 | SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProlys 546 |
| QY | 669 | GCTACACATCAAAAAGAAATGGAATAAATAAGTGGAAAATTAGAAGATTCAACTAGCCTA 728 |
| DB | 547 | AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 566 |
| QY | 729 | TCAAAATCTTGATACAGTTCATCTTGTGAAAGAGCAGGAGGAATCTCAAAAGATCAC 788 |
| DB | 567 | SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 586 |
| QY | 789 | TGTGAACACTACAGGAAAAATGGAAACAAATGAAAGAAAGTTTTTGTGTACTGAAAAAG 848 |
| DB | 587 | CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 606 |
| QY | 849 | AAACTGTCAAGAACAAAGAAATAAATCACAGTTTAGAGAACCAAAAAAGTTAAATGGGAA 908 |
| DB | 607 | LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysTrpGlu 626 |

QY 1149 CATGAAATTCGATGTTGAAAGAAATTCCTCATCTGTAAGCTGGAATAGCCACACTG 1208
 Db 707 HisGluAsnCysMetLeuLysGluLeuAlaMetLeuLysLeuGluLeuAlaThrLeu 726
 QY 1209 AAACACCAATACACGAAAGAAATTAATACCTTTGAGGACATTAAGATTTTAAAGAA 1268
 Db 727 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 746
 QY 1269 AGAATGCTGAATTCAGATGACCTCAAACTGAAAGAGAAATCATTAATCAAAAGGCA 1328
 Db 747 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 766
 QY 1329 TCTCAATATAGTGGCAGCTTAAGCTTCTGATAGCTGAGACACAACTCTCCTCTAAA 1388
 Db 767 SerGlnTyrSerGlnLeuLysValLeuLeuAlaGluAsnThrMetLeuThrSerLys 786
 QY 1389 TTGAAGGAAACAAAGAAATTAATGAGGAGAAATTAATGATCAACCATCTCTAGA 1448
 Db 787 LeuLysGluLysGlnAspLysGluLeuLysGluAlaGluLeuGluSerHisHisProArg 806
 QY 1449 CTGGCTTCTGCTACAGACCATCAATTAATTTGACATCAAGAAAGTCAAGACCT 1508
 Db 807 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 826
 QY 1509 GCTTTCCACATTCAGAGATGCTTTGTTGCAAGAAATCAATTTGATGTGAGTAGT 1568
 Db 827 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 846
 QY 1569 ACCGATATATAAATGAGTCTCCATCAACCATCTTCTGAAGCTCAAGAAATCCAN 1628
 Db 847 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLy 866
 QY 1629 AAGCCTAAATTAATCTCAATTTATGACAGAGATGCTCTAAGAGAAATACATTTGTTTC 1688
 Db 866 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 886
 QY 1689 AGGAACATGCAAAAGAGACCAACCTGAAACACAGTGTCAATGAAGAGACTGAACACA 1748
 Db 886 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 906
 QY 1749 TGTNTCAANGAACAAGATNATGGAACAACACACTGANCAGCAGGAGTCTAGATC 1808
 Db 906 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 926
 QY 1809 AGAAATTTTCACTACAAAGCAAAATATGTGCTTCAACAGCAATTAGTTCATGCAC 1868
 Db 926 InLysLeuPheGlnLeuGlnSerLysAsnMetTyrLeuGlnGlnLeuValHisAlaH 946
 QY 1869 ATAANGAAGCTGACAAAGAGAGATACAAATTTGATNTTCATNTCTTTCAGAGGAA 1928
 Db 946 Is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 965
 QY 1929 ATGCN-CATCATCTTCTAAAAGAGAAATGAGAGATATTTNATTCATNATCAATTTA 1987
 Db 966 MetGlnHisHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 985
 QY 1988 AAAACCCGTATATTCAATATGGAAGAAAAAANAANA 2027
 Db 986 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 998

RESULT 8

ABU37741
 ID ABU37741 standard; protein; 1002 AA.
 AC
 XX
 AC ABU37741;
 XX
 DT 15-MAY-2003 (first entry)
 DE Human tumour-related protein - SEQ ID NO 475.
 XX Human; vaccine; gene therapy; T cell stimulation; T cell expansion;
 KW tumour; breast cancer; cancer; immune response stimulation.

XX Homo sapiens.
 OS WO200283956-A1.
 XX
 XX 24-OCT-2002.
 XX
 XX 15-APR-2002; 2002WO-US012378.
 XX
 PR 13-APR-2001; 2001US-00834759.
 PR 07-DEC-2001; 2001US-00007805.
 PR 13-FEB-2002; 2002US-00076622.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 XX Houghton RL, Sleath PR, Persing DH, Jiang Y, Dillon DC;
 PI Mitcham JL, Xu J, Harlocker SL, Hepler WT, Henderson RA, Fanger GR;
 PI Vedvick TS, McNeill PD, Durham M;
 XX
 DR WPI; 2003-103376/09.
 XX
 PT New polypeptide and polynucleotide useful for stimulating and/or
 PT expanding T cells specific for a tumor protein and treating breast
 PT cancer.
 XX
 PS Example 1; Page 298-300; 375pp; English.
 XX
 CC The invention comprises a method of stimulating and/or expanding T cells
 CC specific for a tumor protein. The invention further comprises human
 CC nucleic acids and proteins that are associated with tumors (e.g. breast
 CC cancer). The method and sequences of the invention are useful for
 CC stimulating and/or expanding T cells specific for a tumor protein,
 CC detecting the presence of cancer, stimulating an immune response in a
 CC patient and treating breast cancer. The present amino acid sequence
 CC represents a human tumour-related protein.
 XX
 SQ Sequence 1002 AA;
 Alignment Scores:
 Pred. No.: 1.66e-273 Length: 1002
 Score: 3261.00 Matches: 654
 Percent Similarity: 97.19% Conservative: 3
 Best Local Similarity: 96.75% Mismatches: 15
 Query Match: 95.83% Indels: 5
 DB: Gaps: 0
 US-09-602-362E-15 (1-2030) x ABU37741 (1-1002)
 QY 9 GTTAAAGATGCTCTCTGAAGGCTAACTGCGAATGAAAGTTTCTATTCCAACTAAAGCC 68
 Db 327 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 346
 QY 69 TTAGAATTGATGACATGCAAACTTTCAAAGCAGAGCCCTCCGAGAGCCATCTGCCTTC 128
 Db 347 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 366
 QY 129 GAGCCTCCCATTCGAATGCAAAAGTCTGTTCCAAATAAAGCCCTTGGAAATGAGAAATGA 188
 Db 367 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuLeuLysAsnGlu 386
 QY 189 CAAACATTTGAGAGCAGATGAGATCTCCCATCAGAAATCCAAACAAAGACATATGAAGAA 248
 Db 387 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 406
 QY 249 AGTTCTTGGATTCAGAGATCTCTGTGAGACTGTTTCAGAGAGATGTGTGTTTACCC 308
 Db 407 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 426
 QY 309 AAGGCTACACATCAAAAGAAATAGATATAAATAAATGGAATAATAGAGAGTCTCTGAT 368
 Db 427 LysAla***HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 446
 QY 369 AATGATGGTTTCTGAAGGCTCCCTGCAGAAATGAAAGTTTCTATTCCAACTAAAGCCTTA 428

Db 447 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 466
QY 429 GAATTGATCGACATGCAAACTTTCAAGACAGAGCCTCCCGAGAAGCCATCTGCCTTCGAG 488
Db 467 GluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu 486
QY 489 CCTGCCATTGAATGCAAAAGTCTGTTCCAAATAAGCCCTTGAATTGAAGNATGAACAA 548
Db 487 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 506
QY 549 ACATTGAGAGCAGATCAGATGTTTCCCTTCAGAAATCAAAACAAAGAAAGTTTGAAGAAAT 608
Db 507 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys**ValGluGluAsn 526
QY 609 TCTTGGGATTCGAGAGTCTCCGTGAGACTGTTTCACAGAAGGATGTGTGTACCCCAAG 668
Db 527 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 546
QY 669 GCTACACATCAAAAGAAATGGAATAAATAAGTGAATAATTAGAAGATTCAACTAGCCCTA 728
Db 547 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 566
QY 729 TCAAAATCTTGATACATTCATTCTTGTGAAAGCAAGGAACTTCACAAAAGATCAC 788
Db 567 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 586
QY 789 TGTGAAACACGTACAGGAAATGGAACAAATGAAAGAAAGTGTGTACTGCTGAAAGAA 848
Db 587 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys 606
QY 849 AAACCTGTCAGAGCAAAAGAAATAAATCACAGTTAGAGAACCAAAAGTTTAAATGGAA 908
Db 607 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 626
QY 909 CAAGACTCTGAGTGTGAGTGTACTTAAACCAAGAGAGAGAGAGAGAGAGATGCC 968
Db 627 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 646
QY 969 GATATATAAATCAAAATTAAGGAAGAAATAGGAAGAAATCGAAGACAGCATAGGAAA 1028
Db 647 AspIleLeuAsnGluLysIleArgGluLeuGlyArgIleGluGluGlnHisArgLys 666
QY 1029 GAGTTAGAGTGAACAACTGAACTGAAAGGCTCTCAGATACAGATATAGAAATGAAG 1088
Db 667 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLys 686
QY 1089 AGTGTAGAAAGTAATTTGAATCAGGTTTCTCACACTCATGAAATGAAATTTATCTCTTA 1148
Db 687 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 706
QY 1149 CATGAAATTCGATGTTGAAAGAAATGGAATGTCATGCTTAAACTGGAATAGCCACATG 1208
Db 707 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 726
QY 1209 AAACACCAATACAGAGAAAGAAATAAATACTTTGAGGACATTAAGATTTTAAAGAA 1268
Db 727 LysHisGlnLysGlnGlnLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 746
QY 1269 AAGAATGCTGAATTCAGATGACCCCTTAAACTGAAAGAGGAATCAATTAACATAAGGCGA 1328
Db 747 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluLysSerLeuThrLysArgAla 766
QY 1329 TCTCAATATAGTGGGAGCTTAAAGTCTGATAGCTGAGAACACATGCTCACTCTTAA 1388
Db 767 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 786
QY 1389 TTGAAGAAAAACAGACAAAGAAATCTAGAGGAGAAATGAATTCACCATCCTAGA 1448
Db 787 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 806
QY 1449 CTGGCTCTCTGTACAGACCATGATCAATGTGCATCAGAGAAAGTCAGAACT 1508

Db 807 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 826
QY 1509 GCTTTCCACATTCGACAGAGATGCTTGTTCGCAAGAAAAATGAATGTTGATGAGTAGT 1568
Db 827 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 846
QY 1569 ACCGATATATAACAAATGAGGTGCTCCATCAACACATTTCTGAAGCTCAAGGAAATCCAN 1628
Db 847 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 866
QY 1629 AAGCCTAAAAATTAATCTCAATTATGACAGAGATGCTTAAGAGAAAAATACATTGGTTTC 1688
Db 866 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 886
QY 1689 AGGAACATGCACAAAGACACCAACGCGAAACACAGTGTCAAATGAAGAAAGCTGAACACA 1748
Db 886 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 906
QY 1749 TGTNTCAANCGAACAAAGATNATGTCAACAAACACTGANCAGCAGGAGTCTCTAGATC 1808
Db 906 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 926
QY 1809 AGAAATTTTCACTACAAAGCAAAAATATGTGGCTTCAACAGCAATTAGTTTCATGCAC 1868
Db 926 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnLeuValHisAlaH 946
QY 1869 ATAAGAAAGCTGCACAAACAAAGCAAGATAACAATTGATNTTCATTNTCTTCAGAGAAA 1928
Db 946 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 965
QY 1929 ATGCN-CATCATCTTCTTAAAGAGAAAAATAGGAGATATTATTCACNATAACCATTTA 1987
Db 966 MetGlnHisLeuLeuLysGluLysAsnGluIlePheAsnTyrAsnAsnHisLeu 985
QY 1988 AAAAACCCGCTATATTCAATATGCAATATGCAAAAAAANAANAANA 2027
Db 986 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 998
RESULT 9
ADE44427
ID ADE44427 standard; protein; 1002 AA.
AC ADE44427;
XX 29-JAN-2004 (first entry)
DI Human breast cancer protein #10.
DE human; breast tumour; cancer; vaccine; T cell stimulator;
KW T cell expander.
XX Homo sapiens.
OS US2003104366-A1.
FN 05-JUN-2003.
PD 17-APR-2000; 2000US-00551621.
PF 28-DEC-1998; 98US-00222575.
PR 02-APR-1999; 99US-00285480.
PR 23-JUN-1999; 99US-00339338.
PR 02-SEP-1999; 99US-00389681.
PR 03-NOV-1999; 99US-00433826.
XX (JIAN/) JIANG Y.
PA (DILL/) DILLON D C.
PA (MITC/) MITCHAM J L.
PA (XUJJ/) XU J.
PA (HARL/) HARLOCKER S L.
XX Jiang Y, Dillon DC, Mitcham JL, Xu J, Harlocker SL;

DR WPI; 2004-020270/02.
 DR N-PSDB; ADE44426.
 XX
 PT Novel isolated polypeptide comprising immunogenic portion of breast tumor
 PT protein or its variant, useful for formulating vaccines for inhibiting
 PT cancer development in a patient.
 XX

PS Example 1; SEQ ID NO 475; 217pp; English.

XX
 CC The invention relates to an isolated polypeptide comprising at least an
 CC immunogenic portion of a breast tumour protein. The polynucleotide, its
 CC polypeptide, its antibody, a pharmaceutical composition comprising the
 CC fusion protein or the polynucleotide encoding it, a vaccine comprising
 CC the fusion protein or the polynucleotide encoding it, an isolated T cell
 CC population comprising T cells specific for a breast tumour protein, and a
 CC method for removing tumour cells from a biological sample is useful for
 CC inhibiting the development of a cancer in a patient. The polypeptide is
 CC useful for stimulating and/or expanding T cells specific for a breast
 CC tumour protein. Stimulating and/or expanding T cells specific for a
 CC breast tumour protein is useful for inhibiting the development of a
 CC cancer in a patient. The method additionally involves the step of cloning
 CC at least one proliferated cell and then administering the cloned T cells
 CC to the patient. The present sequence represents a breast cancer protein.
 XX
 SQ Sequence 1002 AA;

Alignment Scores:
 Pred. No.: 1,66e-273 Length: 1002
 Score: 3261.00 Matches: 654
 Percent Similarity: 97.19% Conservative: 3
 Best Local Similarity: 96.75% Mismatches: 15
 Query Match: 95.83% Indels: 5
 DB: 8 Gaps: 0

US-09-602-362E-15 (1-2030) x ADE44427 (1-1002)

QY 9 GTTAAAGATGGCTCTCTGAGCTAACTCGGAATGAAAGTTCTTATTCCTAACTAAAGCC 68
 DB 327 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 346
 QY 69 TTAAAGTTGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCG 128
 DB 347 LeuGluLeuMetAspMetGlnThrPhenylsAlaGluProGluGluProSerAlaPhe 366
 QY 129 GAGCTGCGCATTAATGAAAGTCTGTTCCTCAAAATAAGCTTGGAAATGAAGATGAA 188
 DB 367 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 386
 QY 189 CAACATTCGAGCAGATGAGATCTCCATCAGATTCCTCAACAAAGAGACTATGAGAA 248
 DB 387 GlnThrLeuArgAlaAspGluLeuLeuProSerGluSerLysGlnLysAspTyrGluGlu 406
 QY 249 AGTTCTTGAGTCTCTGAGAGTCTCTGAGAGTCTCTGAGAGTCTCTGAGAGTCTCTGAG 308
 DB 407 SerSerTyrAspSerGluSerLeuGluThrValSerGlnLysAspValCysLeuPro 426
 QY 309 AAGGCTACATCAAAAGAAATAGATAAATAATGGAATTAAGAGAGTCTCTGAT 368
 DB 427 LysAla***HisGlnLysGluLeuAspLysIleAsnGlyLysLeuGluGluSerProAsp 446
 QY 369 AATGATGTTTCTGAGGCTCCCTCGAGAATGAAAGTTCTTATTCCTCAACTAAAGCCTTA 428
 DB 447 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 466
 QY 429 GAATTGATGAGATGCAAACTTCAAGCAGAGCTCCCGAGAAGCCATCTGCTTCGAG 488
 DB 467 GluLeuMetAspMetGlnThrPhenylsAlaGluProProGluLysProSerAlaPheGlu 486
 QY 489 CTGCGCATTAATGAAAGTCTGTTCCTCAAAATAAGCTTGGAAATGAAGATGAACA 548
 DB 487 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 506
 QY 549 ACATTGAGCAGATCAGATGTTCCCTTCAGAAATCAAAACAAAGAGGTTGAGAAAT 608

DB 507 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys***ValGluGluAsn 526
 QY 609 TCTTGGGATTCTGAGAGTCTCCGTGAGAGTCTTTCACAGAGAGATGTGTGTACCCAAG 668
 DB 527 SerTyrAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 546
 QY 669 GCTACACATCAAAAGAAATGGAATAAATAGTGGAAATTAGAAGATCAATCACTAGCCTA 728
 DB 547 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 566
 QY 729 TCAAAATCTTGATACAGTTCATTCTTGTGAAGAGCAAGGAGACTTCAAAAAGATCAC 788
 DB 567 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 586
 QY 789 TGTGAACACGTACAGAAAAATGGAACAAATGGAACAAATGGAACAAATGGAACAAATGGA 848
 DB 587 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 606
 QY 849 AAATCTGTCAAGCAAAAGAAATATAAATACACAGTATAGAGAACCAAAAGATTAATCGGA 908
 DB 607 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTyrGlu 626
 QY 909 CAAGAGCTGTGAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAGAGAGAGAGAG 968
 DB 627 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 646
 QY 969 GATATATTAAATCAAAATTTAGGAGAGATTTAGGAGAGATTTAGGAGAGATTTAGGAG 1028
 DB 647 AspIleLeuAsnGlnLysIleArgGluLeuGluGlyArgIleGluGlnHisArgLys 666
 QY 1029 GAGTTAGAGTGAACCAACATTTGAACAGGCTCTCAGATAACAAGATATAGATTTGAAG 1088
 DB 667 GluLeuGluValLysGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 686
 QY 1089 AGTGTAGAAATTAATTTGATCAGTGTCTCAGACTCATGAAATGAAATTAATTTCTCTTA 1148
 DB 687 SerValGlnSerAsnLeuAsnGlnValSerHisThrHisGlnAsnGluAsnThrLeu 706
 QY 1149 CATCAAAATTCATGTTGAAAAAGGAAATTCCTCATGCTAAACTGAAATAGCCACTG 1208
 DB 707 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 726
 QY 1209 AACACCAATACAGGAAAGGAAATAAATATCTTTGAGCAGATTAAGATTTAAAGAA 1268
 DB 727 LysHisGlnThrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 746
 QY 1269 AAGAACTCTCAACTTCAGATGACCTTAAACTCAAGAGAGAAATCATTAATAAAGGCA 1328
 DB 747 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluSerLeuThrLysArgAla 766
 QY 1329 TCTCAATATAGTGGCAGCTTAAAGTTCTGATGCTGAGAACACATGCTCCTCTTAA 1388
 DB 767 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 786
 QY 1389 TTGAGGAAAAACAAGCAAGAAATACTAGAGCAAGAAATTCATCAACACACCTCTCTAGA 1448
 DB 787 LeuLysGlnLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisProArg 806
 QY 1449 CTGCTCTCTGCTGTACAGACCATGATCAAAATTTGACATCAAGAAAGAACTCAAGAACCT 1508
 DB 807 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 826
 QY 1509 GCTTTCACATTCGAGAGATGCTTGTTCGCAAGAAATTAATGATGTTGATGTGAGTAGT 1568
 DB 827 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 846
 QY 1569 ACCGATATATAACAAATGAGGTGCTCCATCAACACCTTCTTGAAGCTCAAGAAATCCAN 1628
 DB 847 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 866
 QY 1629 AAGCTTAAATTAATCTCAATTATGACAGAGATGCTCTTGAAGAAATACATTGTTTC 1688

Db 866 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 886
 QY 1689 AGGAACATGCACAAAGAGACCAAGTGAACACAGAGTCAAAATGAAGAGAGTGAACACA 1748
 Db 886 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 906
 QY 1749 TGTNTCAAACGACAAAGATNTGTGAACAAACACACACTGANCAGCAGGAGTCTCTAGATC 1808
 Db 906 eTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGlnSerLeuAspG 926
 QY 1809 AGAAATATTTCAACTACAAAGCAAAATATGTGGCTTCAACAGCAATAGTTCATGCAC 1868
 Db 926 lnyLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnValHisAlaH 946
 QY 1869 ATAANGAAGCTGACACAAACAAAGACAGATAACAATGATNTTCATNTCTTGAGAGGAAA 1928
 Db 946 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 965
 QY 1929 ATGCN-CATCATTTCTAAAGAGAAAATGAGGAGATATTNATTACNATTAACCATTTA 1987
 Db 966 MetGlnHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 985
 QY 1988 AAAAAACCGTATATTTCAATATGGAAGAAAAAANAARAA 2027
 Db 986 LysAsn-ArgIleYrGlnTyrGluLysGluLysAlaGlu 998

RESULT 10
 AAU33357
 ID AAU33357 standard; protein; 1095 AA.

AC AAU33357;
 DT 18-DEC-2001 (first entry)
 XX Human breast cancer protein B726P fusion protein #1.
 DE Human; ss; breast cancer protein; tumour; cancer; cytostatic;
 KW gene therapy; immunogen.
 KW Homo sapiens.
 OS

XX WO200179286-A2.
 PN 25-OCT-2001.
 XX 12-APR-2001; 2001WO-US012164.
 XX 17-APR-2000; 2000US-00551621.
 PR 08-JUN-2000; 2000US-00590751.
 PR 22-JUN-2000; 2000US-00604287.
 PR 20-JUL-2000; 2000US-00620405.
 XX (CORI-) CORIXA CORP.
 PA
 XX Jiang Y, Dillon DC, Mitcham JL, Xu J, Harlocker SL, Hepler WT;
 PI WPI; 2001-611721/70.
 XX N-PSDB; AAS47421.
 DR
 XX Breast Tumor Proteins and nucleic acids useful for the prevention,
 PT diagnosis and treatment of breast cancer.
 XX Claim 22; Page 292-295; 297pp; English.
 XX The invention relates to isolated breast tumour proteins and nucleic
 CC acids that encode them, including immunogenic fragments of the proteins.
 CC Also included are expression vectors expressing the proteins, transformed
 CC cells and antibodies raised against the proteins or an antigen presenting
 CC cell expressing the protein. The proteins and nucleic acids may be used
 CC in the prevention, diagnosis and treatment of diseases associated with
 CC inappropriate breast tumour protein expression, i.e. breast tumours and
 CC breast cancer e.g by gene therapy. The nucleic acids and their
 CC complements may also be used as DNA probes in diagnostic assays to detect

CC and quantitate the presence of similar nucleic acids in samples, and
 CC therefore which patients may be in need of restorative therapy. The
 CC proteins, nucleic acids and antibodies may be used in assays to identify
 CC modulators (e.g. antagonists) of breast tumour protein expression and
 CC activity. The antibodies and antagonists may also be used to down
 CC regulate expression and activity. The antibodies may also be used as
 CC diagnostic agents for detecting the presence of the proteins in samples
 CC (e.g. by enzyme linked immunosorbant assay (ELISA)) and in other immuno-
 CC purification diagnostic techniques. The present sequence is a breast
 CC tumour protein encoded by a cDNA from a breast tumour cDNA library
 CC isolated by subtractive hybridisation against a normal breast cDNA
 CC library
 XX
 SQ Sequence 1095 AA;
 Alignment Scores:
 Pred. No.: 1,71e-273 Length: 1095
 Score: 3261.00 Matches: 654
 Percent Similarity: 97.19% Conservative: 3
 Best Local Similarity: 96.75% Mismatches: 15
 Query Match: 95.83% Indels: 5
 DB: Gaps: 0
 US-09-602-362E-15 (1-2030) x AAU33357 (1-1095)
 QY 9 GTTAAAGATGGTCTTCTGAAGGCTAACTCGGAGATGAAAGTTTCTATTTCACAACTAAGCC 68
 Db 420 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 439
 QY 69 TTGAGATTGATGACATGCAAACTTTTCAAGCAGAGCCTCCGAGAGGCATCTGCCTTC 128
 Db 440 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 459
 QY 129 GAGCCTGCCATTGAAATGCAAAAGTCTGTTCAAATAAAGCCCTTGAATGAAGATGAA 188
 Db 460 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 479
 QY 189 CAACATTTGAGACAGATGATGATCTCCATCAGATCCAAACAAAGAGCATTATGAAGA 248
 Db 480 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 499
 QY 249 AGTTCTTTGGATTCTCAGAGTCTCTGTGAGACTGTTTTCACAGAGGATGTGTGTTCACC 308
 Db 500 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 519
 QY 309 AAGCTACACATCAAAAAGAAATAGATAAATAATGAATGAAGAGTCTCTGAT 368
 Db 520 LysAla***HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluSerProAsp 539
 QY 369 AATGATGTTTCTGAGAGTCTCCTCGCAATGAAAGTTTCTATTCCAACTAAAGCCTTA 428
 Db 540 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 559
 QY 429 GAATTGATGACATGCAAACTTTCAAGCAGAGCCTCCGAGAGCATCTGCCTTCGAG 488
 Db 560 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 579
 QY 489 CCTGCCATTGAAATGCAAAAGTCTGTTCAAATAAAGCCTTGAATGAAGATGAACAA 548
 Db 580 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 599
 QY 549 ACATTGAGACAGATCAGATGTTCCCTTCAGAAATCAAAAACAAAGAGGTTCAAGAAAT 608
 Db 600 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys***ValGluGluAsn 619
 QY 609 TCTTTGGATTCTCAGAGTCTCCTGTGAGACTGTTTTCACAGAGGATGTGTGTTCACCAAG 668
 Db 620 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 639
 QY 669 GCTACACATCAAAAAGAAATGATAAATAATGTTGAAGAAATTAAGAGATTCACATGACCTA 728
 Db 640 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 659

QY 729 TCAGAAATCTTGATACAGTTCATTCTTGTGGAAGCAAGGAACTTCAAAAGATCAC 788
Db 660 SerLysileLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 679
QY 789 TGTGAACAAACGTACAGCAAAATGGAACAAATGAAAGAAAGTGTGTGTACTGAAAAAG 848
Db 680 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 699
QY 849 AAATCTGAGACAAAGAAATAAATACAGTGTAGAGAACCAAAAGTAAATGGGAA 908
Db 700 LysLeuSerGluAlaLysGluileLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 719
QY 909 CAAGAGCTCGCAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAAATGCC 968
Db 720 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 739
QY 969 GATATATTAATGAAATAATAGGGAAGATTAGGAAGATCGAAGAGCAGCATAGGAAA 1028
Db 740 AspIleLeuAsnGluLysileArgGluGluLeuGlyArgIleGluGluGlnHisArgLys 759
QY 1029 GAGTTAGAGTGAACCAACACTTGAAACAGGCTCTCAGATACAGATACAGATATGAA 1088
Db 760 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 779
QY 1089 AGTGTAGAAATTAATTTGAATCAGGTTCTCTCAGCTCATGAAATGAAATTTATCTCTTA 1148
Db 780 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 799
QY 1149 CATGAAATTCGATTTGAAAGAGAAATGCGATCTCAAACTGGAATAGCCACACTG 1208
Db 800 HisGluAsnCysMetLeuLysLysGluileAlaMetLeuLysLeuGluileAlaThrLeu 819
QY 1209 AAACACCAATACAGGAAAGAAATAATAATCTTTGAGGACATTAAGATTTTAAAGAA 1268
Db 820 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysileLeuLysGlu 839
QY 1269 AGAATGCTGAATCTCAGATGACCCCTAAACCTGAAAGAGGAATCATTAATAAGGGCA 1328
Db 840 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 859
QY 1329 TCTCAATATAGTGGGAGCTTAAAGTTCGTAGTGTGAGACACAACTGCTCACTTCTAAA 1388
Db 860 SerGlnTyrSerGlyGlnLeuLysValLeuileAlaGluAsnThrMetLeuThrSerLys 879
QY 1389 TTGAAGGAAAAACAAGCAAGAAATATAGAGGAGCAAAATTCATACACCATCCTAGA 1448
Db 880 LeuLysGluLysGlnAspLysGluileLeuGluAlaGluileGluSerHisProArg 899
QY 1449 CTGGCTTCTCTACAGACCATCATCAATGTGCACATCAAGAAAGTCAAGAACCT 1508
Db 900 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 919
QY 1509 GCTTTCCACATTCAGAGATGCTTTGTTGCAAGAAAAATGAATGTTGATGTGAGTAGT 1568
Db 920 AlaPheHisileAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 939
QY 1569 ACGGATATATACATAGAGGTGCTCATCAACACTTTCTGAAGCTCAAGGAATCCAN 1628
Db 940 Thr-ileTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerly 959
QY 1629 AAGCCTAAAAATTAATCTCAATATATGACGAGATGCTTAAGAGAAAAATACATGTTTC 1688
Db 959 sSerLeuLysileAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 979
QY 1689 AGGAACATGCACAAAGACCAACGTGAAACACAGTGTCAATGAAGGAAGTGAACACA 1748
Db 979 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 999
QY 1749 TGTTNTCAACAGCAAGATNATGTCAACAAACACTGANCAGCAGGAGTCTTAGATC 1808
Db 999 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1019
QY 1809 AGAAATTTTCAACTACAAAGCAAAATATATGGCTTCAACAGCAATTAGTTTCATGCAC 1868

Db 1019 InLysLeuPheGlnLeuGlnSerLysAsnMetTyrLeuGlnGlnLeuValHisAlaH 1039
QY 1869 ATAANGAAAGCTGCAACAAAGCAAGATAACAAATTCATTTCTTTGAGAGAAA 1928
Db 1039 is-LysLysAlaAspAsnLysSerLysileThrileAspIleHisPheLeuGluArgLys 1058
QY 1929 ATGCN-CATCATCTCTTAAAGAGAAAAATGAGAGATATTTNATTACNATAACCATTA 1987
Db 1059 MetGlnHisLysLeuLeuLysGluLysAsnGluilePheAsnTyrAsnAsnHisLeu 1078
QY 1988 AAAAACCCGATATTTCAATATGCAAAAAAANAANA 2027
Db 1079 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1091
RESULT 11
ABG78924
ID ABG78924 standard; protein; 1095 AA.
XX AC ABG78924;
XX 15-NOV-2002 (first entry)
DT Human breast tumour polypeptide #15.
XX Human breast tumour polypeptide #15.
DE Human; breast tumour protein; breast cancer; cytostatic; vaccine.
XX Human sapiens.
XX OS
XX US2002085998-A1.
XX PN
XX 04-JUL-2002.
XX 13-APR-2001; 2001US-00834759.
XX 28-DEC-1998; 98US-00222575.
XX 02-APR-1999; 99US-00285480.
XX 23-JUN-1999; 99US-00339338.
XX 02-SEP-1999; 99US-00389681.
XX 03-NOV-1999; 99US-00433826.
XX 17-APR-2000; 2000US-00551621.
XX 08-JUN-2000; 2000US-00590751.
XX 22-JUN-2000; 2000US-00604287.
XX 20-JUL-2000; 2000US-00620405.
XX (CORI-) CORIXA CORP.
XX Jiang Y, Dillon DC, Mitcham JL, Xu J, Harlocker SL, Hepler WT;
XX Henderson RA;
XX WPI; 2002-635657/68.
XX N-ESDB; ABS64022.
XX Novel breast cancer polynucleotides and polypeptides encoded by the
XX polynucleotides, useful for detecting the presence of breast cancer in a
XX patient, and in pharmaceutical compositions, for treating breast cancer.
XX Disclosure; Page 223-225; 247pp; English.
XX The invention relates to an isolated breast tumour polynucleotide and the
XX polypeptide it encodes. The polynucleotide and polypeptide are useful for
XX detecting the presence of breast cancer in a patient, and in
XX pharmaceutical compositions for treating breast cancer. The sequences are
XX useful for stimulating an immune response in a patient and can therefore
XX be used in production of vaccines. The sequences are also useful for
XX detecting the presence of a cancer in a patient, by obtaining a
XX biological sample from the patient, contacting the biological sample with
XX a composition of the invention and detecting the amount of polynucleotide
XX that hybridizes to the sample. This sequence represents a human breast
XX tumour polypeptide of the invention
XX Sequence 1095 AA;

Alignment Scores:

Pred. No.: 1,71e-273 Length: 1095
 Score: 3261.00 Matches: 654
 Percent Similarity: 97.19% Conservative: 3
 Best Local Similarity: 96.75% Mismatches: 15
 Query Match: 95.83% Indels: 5
 DB: 5 Gaps: 0

US-09-602-362E-15 (1-2030) x ABG78924 (1-1095)

Qy 9 GTTAAAGATGGTCTCTGAAAGGCTAACTGCGGAATGAAGTTTCTATTTCACAAATGAAGCC 68
 Db 420 ValIysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 439
 Qy 69 TTAGAATTGATGACATGCAAACTTTCAAGACGAGACCTCCCGAGAGCAATCTGCTTC 128
 Db 440 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 459
 Qy 129 GAGCTGCCATTGAAATGCAAAAGTCTGTTCCTCAAAATAAAGCCCTGGAATGAAGATGAA 188
 Db 460 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 479
 Qy 189 CAACATGACAGACATGAGTACTCCCATCAGATCCAAACAAAGAGGACTATGAAGA 248
 Db 480 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 499
 Qy 249 AGTTCTGGGATTCGAGAGTCTCTGTGAGACTGTTTCACAGAGGATGTGTTPACCC 308
 Db 500 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 519
 Qy 309 AAGCTACACATCAAAAGAAATAGATAAATAATGAAATGAAAGTCTCTGTAT 368
 Db 520 LysAla***HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 539
 Qy 369 AATGATGGTTTCTGAAGGCTCCCTGCAGAAATGAAAGTTTCTATTCCAACTAAAGCCCTTA 428
 Db 540 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 559
 Qy 429 GAATGTGATGACATGCAAACTTTCAAGACGAGACCTCCCGAGAGCAATCTGCTTCGAG 488
 Db 560 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 579
 Qy 489 CTTGCCATTCAATGCAAAAGTCTGTTCCTCAAAATAAAGCCCTGGAATGAAGATGAACAA 548
 Db 580 ProIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 599
 Qy 549 ACATTGAGACATGATGTTCCCTTCAGATCAAAACAAAGAGGTTTGAAGAAAT 608
 Db 600 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys***ValGluGluAsn 619
 Qy 609 TCTTGGGATTCGAGATCTCCGTGAGACTGTTTCACAGAGGATGTGTGTACCCCAAG 668
 Db 620 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 639
 Qy 669 GCTACACATCAAAAGAAATGGAATAAATGAATGGAATAATGAAGATCAACTAGCCTA 728
 Db 640 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 659
 Qy 729 TCATAAATCTGGATACAGTCTCCTTGTGAAAGACGAGGAACTCAAAAGATAC 788
 Db 660 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 679
 Qy 789 TGTCAACAACTCAGGAAATGGAACAAATGAAAGAAAGTTTGTGTACTGAAAGAA 848
 Db 680 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys 699
 Qy 849 AAATGTTCAGACGAAAGAAATAAATACAGTTAGACCAACAAAGTAAATCGGAA 908
 Db 700 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 719
 Qy 909 CAAGAGCTCTGAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAGAAATGCC 968
 Db 720 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 739

Qy 969 CATATATTAAATGAAAAAATTAGGGAAGAATTAGGAAGAATCGAAGACGACATAGGAAA 1028
 Db 740 AspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGluGlnHisArgLys 759
 Qy 1029 GAGTTAGAGTGAACCAACACTTGAACAGGCTCTCAGATACAGATATAGATTAAGTGAAG 1088
 Db 760 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 779
 Qy 1089 AGTGTAGAAATAATTTGAATCAGTTTCTCACACTCATGAAATGAAATATCTCTTA 1148
 Db 780 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 799
 Qy 1149 CATGAAATTTGCATGTTCAAAAAAGGAAATGCCATGCTAAACTGGAATATAGCCACACTG 1208
 Db 800 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 819
 Qy 1209 AAACACCAATACCAAGGAAAGGAAATAAATACTTTGAGGACATTAAGATTTTAAAGAA 1268
 Db 820 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 839
 Qy 1269 AAGATGCTGAACCTTCAGATGACCTAAAACTGAAAGAGGAATCATTAATAAAGGCA 1328
 Db 840 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 859
 Qy 1329 TCTCAATATAGTGGGAGCTTAAAGTTCTGATAGCTGAGACACACATGCTCACTCTAAA 1388
 Db 860 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 879
 Qy 1389 TTGAAGGAAAAACAAGCAAGAAATACTAGAGGACAGAAATGAATCACACCATCTCTAGA 1448
 Db 880 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluLeuGluSerHisProArg 899
 Qy 1449 CTGCTCTCTGTGTACAAAGCCATGATCAAAATGTGTGACATCAAGAAAGTCAAGAACCT 1508
 Db 900 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 919
 Qy 1509 GCTTTCCACATTCGAGGAGATGCTTGTTCGAAAGAAAAATGAATGTGTGATGTGAGTAGT 1568
 Db 920 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 939
 Qy 1569 ACCGATATTAACAAATGAGGTGCTCCATCAACCACTTCTGAGCTCAAGGAAATCCAN 1628
 Db 940 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 959
 Qy 1629 AAGCTAAAAATTAATCTCAATTATGAGGAGATGCTCTAAGAGAAATACATTGGTTTC 1688
 Db 959 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 979
 Qy 1689 AGGAACATGCAAAAGACCAACGTCGAAACACACAGTCTCAAAATGAAGAGAGCTGAACACA 1748
 Db 979 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHis 999
 Qy 1749 TGTNTCAANCGACACAGATNATCTGAACAAACACACTGANCAGACGAGTCTCTAGATC 1808
 Db 999 eTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGlnLysSerLeuAspG 1019
 Qy 1809 AGAAATTTTCAACTACAAAGCAAAATATGTGGCTTCAACAGCAATTAAGTTTCATGCAC 1868
 Db 1019 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 1039
 Qy 1869 ATANGAAAGCTGACACAAAGCAAGCAATTAACATTTGATTTTCATTTCTTGAGAGGAAA 1928
 Db 1039 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1058
 Qy 1929 ATGCTCATCTCTTAAAAAGAAAAATGAGGAGATTTTATTTATTTATTTATTTATTTA 1987
 Db 1059 MetGlnHisHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1078
 Qy 1988 AAAAACCCTGATTTTCAATGGAAGAAAAAANAANAANAANAANAANAANAANAANAANA 2027
 Db 1079 LysAsn-ArgIleTyrGlnTyrGlnLysGluLysGluLysAlaGlu 1091

RESULT 12
ABJ37747
ID ABJ37747 standard; protein; 1095 AA.
XX
XX AC
XX ABJ37747;
DT 15-MAY-2003 (first entry)
XX
DE Human tumour-related protein - SEQ ID NO 493.
XX
XX Human; vaccine; gene therapy; T cell stimulation; T cell expansion;
XX tumour; breast cancer; cancer; immune response stimulation.
XX
XX Homo sapiens.
XX
XX WO200283956-A1.
XX
XX PD 24-OCT-2002.
XX
XX 15-APR-2002; 2002WO-US012378.
XX
XX 13-APR-2001; 2001US-00834759.
XX
XX 07-DEC-2001; 2001US-00007805.
XX
XX 13-FEB-2002; 2002US-00076622.
XX
XX (CORI-) CORIXA CORP.
XX
XX Houghton RL, Sleath PR, Persing DH, Jiang Y, Dillon DC;
XX Mitcham JL, Xu J, Harlocker SL, Hepler WT, Henderson RA, Fanger GR;
XX Vedvick TS, McNeill PD, Durham M;
XX
XX WPI; 2003-103376/09.
XX
XX New polypeptide and polynucleotide useful for stimulating and/or
XX expanding T cells specific for a tumor protein and treating breast
XX cancer.
XX
XX Disclosure; Page 310-312; 375pp; English.
XX
XX The invention comprises a method of stimulating and/or expanding T cells
XX specific for a tumour protein. The invention further comprises human
XX nucleic acids and proteins that are associated with tumours (e.g. breast
XX cancer). The method and sequences of the invention are useful for
XX stimulating and/or expanding T cells specific for a tumour protein.
XX detecting the presence of cancer, stimulating an immune response in a
XX patient and treating breast cancer. The present amino acid sequence
XX represents a human tumour-related protein
XX
XX Sequence 1095 AA;
XX
Alignment Scores:
Pred. No.: 1.71e-273 Length: 1095
Score: 3261.00 Matches: 654
Percent Similarity: 97.19% Conservative: 3
Best Local Similarity: 96.75% Mismatches: 15
Query Match: 95.83% Indels: 5
DB: 6 Gaps: 0
US-09-602-362E-15 (1-2030) x ABJ37747 (1-1095)
QY 9 GTTAAAGATGGTCTTCTGAGGCTAACTCGGAATGAAAGTTCTATTCCAACTAAAGCC 68
Db 420 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 439
QY 69 TTGAATTGATGACATGCAACTTTCAAGCAGAGCCCTCCGAGAGCCATCGCCTTC 128
Db 440 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 459
QY 129 GAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCCTTGGAAATGAAGATGAA 188
Db 460 GluProAlaIleGluMetGlnThrPheLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 479
QY 189 CAAACATTGAGCAGATGATGATCTCCCATCGAATTCCAACAAAGGACTATGAGAA 248

Db 480 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 499
QY 249 AGTCTCTGGATTCCTGAGAGTCTCTGAGACTGTTTTCACAGAGATGTGTGTTTACCC 308
Db 500 SerSerTriPaspSerGluSerLysGluThrValSerGlnLysAspValCysLeuPro 519
QY 309 AAGCTACACATCAAAAAGAAATAGATAAAATAAATGAAAAATTAGAGAGTCTCTGTAT 368
Db 520 LysAla**HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 539
QY 369 AATGATGTTTCTGAAGCTCCCTGCAGAAATGAAAGTTTCTATTCCAACTAAAGCCTTA 428
Db 540 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 559
QY 429 GAATTGATGACATGCAAACTTTCAAAGCAGAGCCCTCCGAGAGCCATCTGCTTCGAG 488
Db 560 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 579
QY 489 CTGCGCAATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGAAGTGAAGTGAACAA 548
Db 580 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 599
QY 549 ACATTGAGAGCAGATCAGATGTTCCCTTCAGAAATCAAAACAAAGAGGTTGAAGAAAT 608
Db 600 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys**ValGluGluAsn 619
QY 609 TCTTGGATTCAGAGTCTCCCTGAGACTGTTTTCAGAGAGATGTGTGTACCCAG 668
Db 620 SerTriPaspSerGluSerLysGluThrValSerGlnLysAspValCysValProLys 639
QY 669 GCTACACATCAAAAAGAAATGGAATAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 728
Db 640 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 659
QY 729 TCAAAATCTTGGATACAGTTCATTCTTGTGAAGAGCAAGCAACTTCAAAAGATCAC 788
Db 660 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 679
QY 789 TGTGAACACGTACAGAAATGGAACAAATGGAACAAATGGAACAAATGGAACAAATGGA 848
Db 680 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys 699
QY 849 AAATCTGTCAGAACAAAGAAATAAATACAGTTAGAGAACCAAAAGTAAATGCGAA 908
Db 700 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 719
QY 909 CAAGAGCTCTGAGTGTGAGATTCATTTAAACCAAGAACAGAGAGAGAGAGAGAGAGCC 968
Db 720 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 739
QY 969 GATATATTAAATGAAAAATTAGGGAAGAAATTAGGGAAGAAATCGAAGAGCAGCATAGAAA 1028
Db 740 AspileLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGluGlnHisArgLys 759
QY 1029 GAGTTAGAAGTGAACCAACAACTTGAACAGCTCTCAGATACAGATATAGATTTGAAG 1088
Db 760 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 779
QY 1089 AGTGTAGAAATGAAATTTGAATTCAGTTCACACTCATCAAAATGAAATTTATCTCTTA 1148
Db 780 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 799
QY 1149 CATCAAAATTCATGTTTGAAGAAAGAAATTCGCAATGCTAAACTGGAATAGCCACTG 1208
Db 800 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 819
QY 1209 AAACACCAATACAGAAAGAAAGAAATTAATACTTTGAGCAGCATTAAGATTTAAAGAA 1268
Db 820 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 839
QY 1269 AAGATGCTGAACCTTCAGATGACCTTAAACTGAAAGAGGAATCATTAACATAAAGGCA 1328

```
Db      840 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluSerLeuThrLysArgAla 859
Qy      1329 TCCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTCAGAACACAAATGCTCCTCTAAA 1388
Db      860 SerGlnTyrSerGlyGlnLeuLysValLeuAlaGluAsnThrMetLeuThrSerLys 879
Qy      1389 TTGAAGGAAAAACAAGACAAGAAATACTAGAGGCGAAATTCGAATCACCACATCTCTAGA 1448
Db      880 LeuLysGlnLysGlnAspLysGluLeuLeuGluAlaGluLeuSerHisProArg 899
Qy      1449 CTGGCTTCTCTCTACAGACCATGATCAAAATGTGCATCAAGAAAAAGTCAAGAACCT 1508
Db      900 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 919
Qy      1509 GCTTTCACATTCGAGAGAGATGCTTGTTCGAAAGAAAAATGAATGCTGATGTAGTAGT 1568
Db      920 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 939
Qy      1569 ACCGATATATAACAATGAGGTGCTCCATCAACCACTTCTGAAGCTCAAGGAAATCCAN 1628
Db      940 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLy 959
Qy      1629 AAGCTTAAATTAATCTCAATATATGACGAGATGCTCTAAGAGAAAAATACATTTGTTTC 1688
Db      959 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgLysThrLeuValSe 979
Qy      1689 AGGAACATGCACAAAGAGACCAACGTGAACACAGTGTCAATGAAGGAGCTGACACA 1748
Db      979 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 999
Qy      1749 TGTNTCAAACNGACAAGATNATGTGAACAACACACTGANCAGCAGGAGTCTCTAGATC 1808
Db      999 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGlnSerLeuAspG 1019
Qy      1809 AGAAATATTTCACACTACAAGCAAAATATGTGGCTTCAACAGCAATTAGTTCATGAC 1868
Db      1019 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 1039
Qy      1869 ATAAGAAAGCTGACACAAACAAAGCAAGATAACAATGATNTTCATTNTCTTGAGAGGAAA 1928
Db      1039 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1058
Qy      1929 ATGCN-CATCATCTTCTAAAGAGAAAATAGGAGATATTTNATACNATACCATTTA 1987
Db      1059 MetGlnHisHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1078
Qy      1988 AAAAAACCGTATATTTCAATATATGGAAGAAAAAANAANA 2027
Db      1079 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1091

RESULT 13
ABJ37783
ID      ABJ37783 standard; protein; 1013 AA.
XX      AC      ABJ37783;
XX      DT      15-MAY-2003 (first entry)
XX      DE      Human tumour-related protein - SEQ ID No 553.
XX      KW      Human; vaccine; gene therapy; T cell stimulation; T cell expansion;
XX      KW      tumour; breast cancer; cancer; immune response stimulation.
XX      OS      Homo sapiens.
XX      PN      WO200283956-A1.
XX      PD      24-OCT-2002.
XX      PF      15-APR-2002; 2002WO-US012378.
XX      PR      13-APR-2001; 2001US-00834759.
XX      PR      07-DEC-2001; 2001US-00007805.
```

```
PR      13-FEB-2002; 2002US-00076622.
XX      (CORI-) CORIXA CORP.
XX      Houghton RL, Sleath PR, Persing DH, Jiang Y, Dillon DC;
PI      Mitcham JL, Xu J, Harlocker SL, Hepler WT, Henderson RA, Fanger GR;
PI      Vedvick TS, McNeill PD, Durham M;
XX      WPI; 2003-103376/09.
XX      New polypeptide and polynucleotide useful for stimulating and/or
PT      expanding T cells specific for a tumor protein and treating breast
PT      cancer.
XX      Example 8; Page 335-338; 375pp; English.
XX      The invention comprises a method of stimulating and/or expanding T cells
CC      specific for a tumour protein. The invention further comprises human
CC      nucleic acids and proteins that are associated with tumours (e.g. breast
CC      cancer). The method and sequences of the invention are useful for
CC      stimulating and/or expanding T cells specific for a tumour protein, a
CC      detecting the presence of cancer, stimulating an immune response in a
CC      patient and treating breast cancer. The present amino acid sequence
CC      represents a human tumour-related protein
XX      Sequence 1013 AA;
SQ      Alignment Scores:
Pred. No.: 5,51e-273 Length: 1013
Score: 3255.00 Matches: 652
Percent Similarity: 97.19% Conservative: 4
Best Local Similarity: 96.59% Mismatches: 15
Query Match: 95.65% Indels: 5
DB: Gaps: 0
```

```
US-09-602-362B-15 (1-2030) x ABJ37783 (1-1013)
Qy      12 AAAGATGCTTCTGAGGCTAACTGCGAATGAAAGTTTCTATTCCAACTAAAGCTTAA 71
Db      339 LysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAlaLeu 358
Qy      72 GAATTGATGACATGCAAACTTTCAAGCAGAGCCCTCCGAGAGGCAATCGCTTCGAG 131
Db      359 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 378
Qy      132 CCTGCCATTCAAAATGCAAAAGTCTGTTCCAAAATAAAGCCCTTGAATGAAGATGAACAA 191
Db      379 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 398
Qy      192 ACATTGAGACAGATGAGATACCTCCCATCGAATCCAAACAAAGCACTATGAGAAAGT 251
Db      399 ThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGluSer 418
Qy      252 TCTTGGGATTCTCAGAGTCTCTGTGAGACTGTGTTTCACAGAAGGATGTGTGTTTACCCTAA 311
Db      419 SerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLys 438
Qy      312 GCTACACATCAAAAAGAAATAGATAAAATAAATGGAATAATTAGAAGAGTCTCTGATAAT 371
Db      439 AlaAlaHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAspAsn 458
Qy      372 GATGGTTTCTGAGGCTCCCTCGACAATCAAGTTTCTATTCCCACTAAAGCCCTTAGAA 431
Db      459 AspGlyPheLeuLysSerProCysArgMetLysValSerIleProThrLysAlaLeuGlu 478
Qy      432 TTGATGACATGCAAACTTTTCAAAGCAGAGCTCCCGAGAGCCATCTCCCTCGAGCCT 491
Db      479 LeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGluPro 498
Qy      492 GCCATTCAATGCAAAAGTCTGTTCCAAATAAGCCTTGAATGAAGATGAACAAACA 551
Db      499 AlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGlnThr 518
```

QY 552 TTGAGACGATCAGATGTTCCCTTCAGAAATCAAAAGAGAGGTTGAGAAAATCT 611
 Db 519 LeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysAsnValGluGluAsnSer 538
 QY 612 TGGGATCTCAGAGTCCGTCGAGCTGTTTCACAGAGGATGTTGTCACCCAGGCT 671
 Db 539 TrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLysAla 558
 QY 672 ACACATCAAAAAGAAATGATAAAATAGTGGAATTTAGAAGATTTCACTAGCCATCA 731
 Db 559 ThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeuSer 578
 QY 732 AAAATCTCGATACAGTCTCATCTCTGTGAAAGAGCAAGGAACTTCAAAAGATCACTGT 791
 Db 579 LysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCys 598
 QY 792 GAACAACTGACAGGAAAAATGAACAAATGAAAGAAAGTGTGTGCTACTGAAAAAGAAA 851
 Db 599 GluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLysLys 618
 QY 852 CTGTGAGAGCAAGAAATAAATCACAGTTAGAGAACCAAAAAGTTAAATGGGAACA 911
 Db 619 LeuSerGluAlaLysGluLysSerGlnLeuGluAsnGlnLysValLysTrpGluGln 638
 QY 912 GAGCTCTGAGTGTGAGATGTTCTTAAACCAAGAGAGAGAGAAATGCCGAT 971
 Db 639 GluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAlaAsp 658
 QY 972 ATATTAAATGAAATAATAGGGAAGAAATPAGGAAGATCGAAGACAGCATAGGAAGAG 1031
 Db 659 IleLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGluGlnHisArgLysGlu 678
 QY 1032 TTAGAAGTGAACAACTTGAACAGGCTCTCAGATACAGATATAGAAATGAGAGT 1091
 Db 679 LeuGluValLysGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLysSer 698
 QY 1092 GTAGAAAATGAAATTTGAAATCAGGTTCTCAGCTCATGAAATGAAATTTATCTCTACAT 1151
 Db 699 ValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeuHis 718
 QY 1152 GAAATTTGATGTTGAAAGAAATGCAATGCTTAAACTGGAATAGCCACATGAAA 1211
 Db 719 GluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeuLys 738
 QY 1212 CACCAATACAGGAAAGAAATAATAATCTTTGAGGACATTAAGATTTTAAAGAAAAG 1271
 Db 739 HisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGluLys 758
 QY 1272 AATGCTGAATCAGATGACCTTAAACTGAAAGAGGAATCAATTAACATAAAGGGCATCT 1331
 Db 759 AsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAlaSer 778
 QY 1332 CAATATAGTGGCAGCTTAAATCTGATAGTGAGACACAACTGCTCACTCTTAATG 1391
 Db 779 GlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLysLeu 798
 QY 1392 AAGGAAAAACAAGAAATACTAGAGGAGAAATGAAATCAACCATCTAGATGAGTACC 1451
 Db 799 LysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisProArgLeu 818
 QY 1452 GCTCTGCTGTACAGACCATGATCAATGTGACATCAAGAAAGTCAAGAACTGCT 1511
 Db 819 AlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluProAla 838
 QY 1512 TTCACATTCGAGGAGATGCTGTTTCAAGAAAAATCAATGTTGATGTAGTAGTACC 1571
 Db 839 PheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerThr 858
 QY 1572 GATATATACATGAGGTGCTCATCAACACATTTCTGAGCTCAAGGAAATCCANAAG 1631
 Db 859 -IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 878
 QY 1632 CCTAAAAATTAATCTCAATATGACGAGATGCTCTTAAGAGAAATACATGTTTCAGG 1691

Db 878 rLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSer-G 898
 QY 1692 AACATGCACAAAGAGACCAACGTGAAACACAGTGTCAAATGAAGAGCTGAACACATGT 1751
 Db 898 luHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisMet 918
 QY 1752 NTCAANCGAACAGATNATGTGAACAAACACACTGANACAGCAGGAGTCTCTAGATCAGA 1811
 Db 918 yrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGlnSerLeuAspGlnL 938
 QY 1812 AATTATTTCACACTACAAAGCAAAATATGTGGCTTCAACAGCAATAGTTTCATGCACATA 1871
 Db 938 ysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaHis- 957
 QY 1872 ANGAAGCTGACCAACAAAGCAAAATCAATTCATNTTCTTGTGAGAGAAAATG 1931
 Db 958 LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluA:GlysMet 977
 QY 1932 CN-CATCATCTTCTAAAGAGAAAATGAGAGATATTTNATTACNATAACCATTTTAAA 1990
 Db 978 GlnHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeuLys 997
 QY 1991 AACCCGTATATTTCAATATGCAAAAAAANAANAANA 2027
 Db 998 Asn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1009
 RESULT 14
 AAB84701
 ID AAB84701 standard; protein; 512 AA.
 XX
 AC AAB84701;
 DT 17-SEP-2001 (first entry)
 XX
 DE Amino acid sequence of a human cancer associated antigen.
 XX
 KW Cancer associated antigen; ING1; tumour suppressor; cancer; vaccine.
 OS Homo sapiens.
 XX
 FN WO2001.47959-A2.
 XX
 PD 05-JUL-2001.
 XX
 PF 29-NOV-2000; 2000WO-US042334.
 XX
 PR 30-NOV-1999; 99US-00451739.
 PR 24-OCT-2000; 2000US-00602362.
 XX
 PA (LUDW-) LUDWIG INST CANCER RES.
 PA (SLOK) SLOAN KETTERING INST CANCER RES.
 PA (CORR) CORNELL RES FOUND INC.
 XX
 PI Jager D, Stockert E, Scanlan M, Knuth A, Old L, Gure A, Chen Y;
 XX
 DR WPI; 2001-441706/47.
 DR N-PSDB; AAH28489.
 XX
 PT Isolated cancer associated nucleic acid molecule identified by SEREX
 PT (serological identification of antigens by recombinant expression
 PT cloning) technique, useful in nucleic acid based therapies to treat
 PT cancer.
 XX
 PS Claim 83; Page 50-51; 62pp; English.
 XX
 CC The present sequence represents a human cancer associated antigen. The
 CC gene is a tumour suppressor candidate gene. The cancer associated antigen
 CC polynucleotides and polypeptides are useful for screening for the
 CC possible presence of a pathological condition in a subject such as
 CC cancer. The cancer associated antigen polypeptides are useful for
 CC producing vaccines

[illegible]

| | | | |
|----------|----------|---|------|
| QY | 942 | CAAGAAGACGACGACGAAGAAATTCGCATATATTAAATCAAAAAAATTCGGGAGAGATTA | 100 |
| Dd | 301 | GlnGluGluGluLysArgAsnAlaAspIleLeuAsnGlnLysIleArgGluGluLeu | 320 |
| QY | 1002 | GGAGAATCGAAGACGACGACGTAGAAAAGAGTTAGACTGAAACAACAACCTTGACAGGCT | 1061 |
| Dd | 321 | GlyArgIleGluGluGlnHisArgLysGluLeuGluValLysGlnGlnLeuGluGlnAla | 340 |
| QY | 1062 | CTCAGATAACAGNATATAGAAATTCGAGAGTGTAGAAAGTAATTTGAATCAGGTTCTCAC | 1121 |
| Dd | 341 | LeuArgIleGlnAspIleGluLeuLysSerValGluSerAsnLeuAsnGlnValSerHis | 360 |
| QY | 1122 | ACTCATGAAATGAAAAATTATCTCTTACATGAAAATTGCATGTTGAAAAAGGAAATGCC | 1181 |
| Dd | 361 | ThrHisGluAsnGluAsnTyrLeuLeuHisGluAsnCysMetLeuLysLysGluIleAla | 380 |
| QY | 1182 | ATGCTAAACTGGAANAATAGCCACACTGAAACCAACATACCAGAAAGGAAAAATAATAC | 1241 |
| Dd | 381 | MetLeuLysLeuGluIleAlaThrLeuLysHisGlnTyrGlnGluLysGluAsnLysTyr | 400 |
| QY | 1242 | TTTTGAGCACAATTAAAGATTTTAAAAAGAAAAGAACTGCTGAACCTTCAGATGACCCCTAAACATG | 1301 |
| Dd | 401 | PheGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuGlnMetThrLeuLysLeu | 420 |
| QY | 1302 | AAAGAGAATCANTAACTAAAGGGGCACTCCAATATAGTCGGCAGCTTAAAGTCTTGATA | 1361 |
| Dd | 421 | LysGluGluSerLeuThrLysArgAlaSerGlnTyrSerGlyGlnLeuLysValLeulle | 440 |
| QY | 1362 | GCTGAGAACCAATGCTCACTTCTTAAATTTGAAGGAAAAACAAGCAAAGAAATCTAGAG | 1421 |
| Dd | 441 | AlaGluAsnThrMetLeuThrSerLysLeuLysGluLysGlnAspLysGluIleLeuGlu | 460 |
| QY | 1422 | GCAGAAATTCGAATCACACCATCTCTAGCTGGCTTCTGCTCTACAAGACCATGATCAAAT | 1481 |
| Dd | 461 | AlaGluIleGluSerHisHisProArgLeuAlaSerAlaValGlnAspHisAspGlnIle | 480 |
| QY | 1482 | GTGCATCAAGAAGAAAGTCAAGAACCTGCTTTCCACATTCAGGAGATGCTGTTTGCAA | 1541 |
| Dd | 481 | ValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAlaCysLeuGln | 500 |
| QY | 1542 | AGAAATGAAATGTTGATGTGCTAGTAGTACCGATATA | 1577 |
| Dd | 501 | ArgLysMetAsnValAspValSerSerThrAspIle | 512 |
| RESULT | 15 | | |
| ABJ37789 | | | |
| ID | ABJ37789 | standard; protein; 1239 AA. | |
| XX | AC | ABJ37789; | |
| XX | DT | 15-MAY-2003 (first entry) | |
| XX | DE | Human tumour-related protein - SEQ ID No 577. | |
| KW | KW | Human; vaccine; gene therapy; T cell stimulation; T cell expansion; | |
| KW | KW | tumour; breast cancer; cancer; immune response stimulation. | |
| OS | OS | Homo sapiens. | |
| XX | FN | WO200283956-A1. | |
| PD | PD | 24-OCT-2002. | |
| PF | PF | 15-APR-2002; 2002WO-US012378. | |
| PR | PR | 13-APR-2001; 2001US-00834759. | |
| PR | PR | 07-DEC-2001; 2001US-00007805. | |
| PR | PR | 13-FEB-2002; 2002US-00076622. | |
| XX | PA | (CORI-) CORIXA CORP. | |
| XX | PI | Houghton RL, Sleath PR, Persing DH, Jiang Y, Dillon DC; | |
| PI | PI | Mitcham JL, Xu J, Harlocker SL, Hepler WT, Henderson RA, Fanger GR; | |


```
Qy 1116 TCTCACACTCATGAAATGAATATCTCTTACATGAAATTCGATGTTGAAAGGAA 1175
Db 918 SerHisThrHisGluSerGluAsnAspLeuPheHisGluAsnCysMetLeuLysLysGlu 937
Qy 1176 ATTGCCATGTAAACATGGAAATAGGCACACTGAAACACCAATACCAGGAAAGGAAAT 1235
Db 938 IleAlaMetLeuLysLeuGluValAlaThrLeuLysHisGlnHisGlnValLysGluAsn 957
Qy 1236 AATACATTTGAGGACATTAAGATTTAAAGAAAGAAATGCTGAACTTCAGATGACCTTA 1295
Db 958 LysTyrPheGluAspIleLysIleLeuGlnGluLysAsnAlaGluLeuGlnMetThrLeu 977
Qy 1296 AACTGAAGAGGAATCATTAACCTAAAGGGCATCTCAATATAGTGGGAGCTTAAAGTT 1355
Db 978 LysLeuLysGlnLysThrValThrLysArgAlaSerGlnTyrArgGlnGlnLeuLysVal 997
Qy 1356 CTGATAGCTGAGAACACATGCTCCTCTTAAATGAAGGAAACCAAGACAAAGAAATA 1415
Db 998 LeuThrAlaGluAsnThrMetLeuThrSerLysLeuLysGluLysGlnAspLysGluIle 1017
Qy 1416 CTAGAGCCAGAAATGATGATCAGACCATCTCTGACTGCTCTGCTGCTACAGACCATGAT 1475
Db 1018 LeuGluThrGluIleGluSerHisHisProArgLeuAlaSerAlaLeuGlnAspHisAsp 1037
Qy 1476 CAAATTTGTGACATCAAGAAAGTCAAGAACCTGCTTCCACATTTGACGAGATGCTTGT 1535
Db 1038 GlnSerValThrSerArgLysAsnGlnGluLeuAlaPheHisSerAlaGlyAspAlaPro 1057
Qy 1536 TTGCAAGAGAAATGATGATGATGAGTAGTACCGATATATAACAATGAGGTGCTCCA 1595
Db 1058 LeuGlnGlyIleMetAsnValAspValSerAsnThr-IleTyrAsnAsnGluValLeuHi 1077
Qy 1596 TCACACACTTTCTGAAGCTCAAGAGAAATCCANAAGCTTAAATAATCTCAATATTATGC 1655
Db 1077 sGlnProLeuTyrGluAlaGlnArgLysSerLysSerProLysIleAsnLeuAsnTyrAl 1097
Qy 1656 AGGAGATGCTCTAAGAGAAATACATTTGTTTTCAGGAACATGACACAAAGAGACCAACGTG 1715
Db 1097 aGlyAspAspLeuArgGluAsnAlaLeuValSer-GluHisAlaGlnA:ga:pa:rg:CysG 1117
Qy 1716 AAACACAGTGTCAAAGAGAGCTGAACACATGTTCAAACGAAACCAAGATNATGTGA 1775
Db 1117 IuThrGlnCysGlnMetClySlyAlaGluHisMetTyrGlnAsnGluGlnAspAsnValA 1137
Qy 1776 ACAACACACTGANCACGAGTCTCTAGATCAGAAATTTATTCAACTACAAAGCAAAA 1835
Db 1137 sPlyHisThrGluGlnGlnGluSerLeuGlnLysLeuPheGlnLeuGluSerLysA 1157
Qy 1836 ATATGTGCTTCAACAGCAATTAAGTTCATGCAATAANGAAGCTGACACACAAAGCAAG 1895
Db 1157 snArgTrpLeuArggGlnGlnLeuValTyrAlaHis-LysLysVal--AsnLysSerLys 1175
Qy 1896 ATAACAATTTGATTTTCATTNTCTTGAGAGGAAATGCNATCATCTTCTTAAAGAGAAAA 1955
Db 1176 ValThrIleAsnIleGlnPheProGluMetLysMet--GlnArgHisLeuLysGluLysA 1195
Qy 1956 ATGAGGAGATATTNATACNATACCATTTAAACCCGATATTTCAATATGCAAAA 2015
Db 1195 snGluGluValPheAsnTyrGlyAsnHisLeuLysGlu-ArgIleAspGlnTyrGluLys 1214
Qy 2016 AAAAAAANAANAANA 2030
Db 1215 GluLysAlaGluArg 1219
```

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 08:46:47 ; Search time 95.3178 Seconds
 (without alignments)
 13439.296 Million cell updates/sec

Title: US-09-602-362E-15
 Perfect score: 3403
 Sequence: 1 ctgcggcgttaagatggt.....gaaaaaaaaaaaaaaaaa 2030

Scoring table: BLOSOM62
 Xgapop 10.0 , Xgapext 0.5
 Ygapop 10.0 , Ygapext 0.5
 Fgapop 6.0 , Fgapext 7.0
 Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Command line parameters:
 -MODEL=frame+ n2p.model -DEV=xlp
 -Q/cgt2.1/USPTO_pool_p/US09602362/runat_15072004_093626_22001/app_query.fasta_1.10325
 -DB=SPREMBL_25 -QFMT=Fascan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
 -UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdi -LIST=45
 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09602362 @CGN 1 1 814 @runat 15072004_093626_22001 -NCPU=6 -ICPU=3
 -NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
 -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPREMBL_25: *
 1: sp_archaea: *
 2: sp_bacteria: *
 3: sp_fungi: *
 4: sp_human: *
 5: sp_invertebrate: *
 6: sp_mammal: *
 7: sp_mhc: *
 8: sp_organelle: *
 9: sp_phase: *
 10: sp_plant: *
 11: sp_rodent: *
 12: sp_virus: *
 13: sp_vertibrate: *
 14: sp_unclassified: *
 15: sp_rvirus: *
 16: sp_bacteriap: *
 17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 3272 | 96.2 | 1341 | 4 Q9BXX3 | Q9bxx3 homo sapien |

| ID | Q9BXX3 | PRELIMINARY; | PRT; | 1341 AA. |
|----|--------|--------------|------|-----------|
| 2 | 1559.5 | 45.8 | 424 | 4 Q9NSI9 |
| 3 | 1546 | 45.4 | 1011 | 4 Q9BXX2 |
| 4 | 1190.5 | 35.0 | 1709 | 4 Q9UFS8 |
| 5 | 1188 | 34.9 | 1710 | 4 Q9HIQ1 |
| 6 | 790.5 | 23.2 | 718 | 4 Q9HCD1 |
| 7 | 671.5 | 19.7 | 823 | 4 Q9H0H6 |
| 8 | 665.5 | 19.5 | 641 | 4 Q60311 |
| 9 | 621 | 18.2 | 453 | 4 Q8IZM7 |
| 10 | 443.5 | 13.0 | 996 | 4 Q8IVF6 |
| 11 | 330 | 9.7 | 2055 | 5 Q8TSC7 |
| 12 | 330 | 9.7 | 2055 | 5 Q8IHP3 |
| 13 | 318.5 | 9.4 | 1738 | 5 Q76329 |
| 14 | 314 | 9.2 | 520 | 11 Q80W27 |
| 15 | 302.5 | 8.9 | 381 | 4 Q15694 |
| 16 | 285 | 8.4 | 2011 | 5 Q8MMC2 |
| 17 | 285 | 8.4 | 2056 | 5 Q9W0W8 |
| 18 | 280 | 8.2 | 2017 | 5 Q94992 |
| 19 | 280 | 8.2 | 2057 | 5 Q94987 |
| 20 | 272 | 8.0 | 1305 | 10 Q9FJ35 |
| 21 | 269 | 7.9 | 2033 | 10 Q7XEH4 |
| 22 | 267 | 7.8 | 1979 | 5 Q96133 |
| 23 | 267 | 7.8 | 2954 | 13 Q42263 |
| 24 | 284.5 | 7.8 | 1313 | 4 Q9HCK7 |
| 25 | 284.5 | 7.8 | 1831 | 10 Q7XN11 |
| 26 | 261 | 7.7 | 1388 | 6 Q28021 |
| 27 | 261 | 7.7 | 2007 | 13 Q02015 |
| 28 | 257.5 | 7.6 | 1320 | 11 Q9JK25 |
| 29 | 257 | 7.6 | 1208 | 5 Q9VXU1 |
| 30 | 257 | 7.6 | 1398 | 5 Q9VXU2 |
| 31 | 256.5 | 7.5 | 895 | 11 Q81112 |
| 32 | 256.5 | 7.5 | 1790 | 3 Q07380 |
| 33 | 256.5 | 7.5 | 1960 | 11 Q8VDD5 |
| 34 | 255.5 | 7.5 | 1762 | 10 Q94DC2 |
| 35 | 255 | 7.5 | 1938 | 6 Q9GJP9 |
| 36 | 255 | 7.5 | 2101 | 4 Q14981 |
| 37 | 255 | 7.5 | 2115 | 4 Q14980 |
| 38 | 254 | 7.5 | 2245 | 5 Q86A36 |
| 39 | 253 | 7.4 | 1416 | 4 Q9BZF9 |
| 40 | 253 | 7.4 | 1416 | 4 Q9HCL1 |
| 41 | 253 | 7.4 | 1449 | 6 Q9BG87 |
| 42 | 253 | 7.4 | 2253 | 13 P70012 |
| 43 | 252 | 7.4 | 947 | 5 Q8ICA7 |
| 44 | 252 | 7.4 | 1388 | 6 Q9GL21 |
| 45 | 251 | 7.4 | 2029 | 4 Q9C014 |

ALIGNMENTS

RESULT 1

Q9BXX3 ID Q9BXX3 PRELIMINARY; PRT; 1341 AA.
 AC Q9BXX3;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Breast cancer antigen NY-BR-1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21174979; PubMed=11280766;
 RA Jaeger D., Stockert E., Guere A.O., Scanlan M.J., Karch J.,
 RA Jaeger E., Knuth A., Old L.J., Chen Y.T.;
 RT "Identification of a tissue-specific Putative Transcription Factor in
 RT Breast Tissue by Serological Screening of a Breast Cancer Library.";
 RL Cancer Res. 61:2055-2061(2001).
 DR EMBL; AF269087; AA27325.1; -;
 DR GO; GO:0005634; C:nucleus; NAS.
 DR GO; GO:0005515; F:protein binding; NAS.
 DR GO; GO:0003700; F:transcription factor activity; NAS.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.

```
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001969; Aspprotease_AS.
DR Pfam; PF00023; ank; 6.
DR SMART; SM00248; ANK; 6.
DR PROSITE; PS00088; ANK REPEAT; 4.
DR PROSITE; PS00297; ANK REP REGION; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR KW ANK repeat; Repeat.
DR ANK repeat; Repeat.
SQ SEQUENCE 1341 AA; 152777 MW; 33E53DDBE6FD3A58B CRC64;

Alignment Scores:
Pred. No.: 9,34e-192 Length: 1341
Score: 3272.00 Matches: 656
Percent Similarity: 97.49% Conservative: 3
Best Local Similarity: 97.04% Mismatches: 13
Query Match: 96.15% Indels: 5
DB: 4 Gaps: 0

US-09-602-362E-15 (1-2030) x Q9BXK3 (1-1341)
QY 9 GTTAAAGATGGTCTTCTGAAGGCTAACTCGGAATGAAAGTTTCTATTCTCAACTAAAGCC 68
DB 666 VAllyAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 685
QY 69 TTGAATTTGATGACATGCAAACTTTCAAGACGAGCCTCCGAGAGCCATCTCGCTTC 128
DB 686 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPhe 705
QY 129 GAGCTGCCATTCAATGCAAAAGTCTGTTCCAAATGAAAGCCTTGGAAATGAAGAAATGAA 188
DB 706 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 725
QY 189 CAAACATTGACGACATGACATCTCCCATCAGAAATCCAAACAAAGACATATGAAGAA 248
DB 726 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 745
QY 249 AGTTCTTGGATCTCAGAGTCTCTGTGAGACTGTTTCAGAAGGATGTGTTTACCC 308
DB 746 SerSerTyrAspSerGluSerIleuCysGluThrValSerGlnLysAspValCysLeuPro 765
QY 309 AAGCTACACATCAAAAGAAATAGATAAAATAATATGAAATTAAGAGAGTCTCTGAT 368
DB 766 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 785
QY 369 AATGATGTTTCTGAAGCTCCCTCGAGAATCAAAAGTTTCTATCCCACTAAAGCCTTA 428
DB 786 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 805
QY 429 GAATTGATGACATGCAAACTTTCAAGCAGAGCCTCCCGAGAAGCCATCTCGCTTCGAG 488
DB 806 GluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu 825
QY 489 CTGCGCATTTGAATGCAAAAGTCTGTTCCAAATAAGACCTTGGAAATGAGAAATGAACA 548
DB 826 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 845
QY 549 ACATTGAGACGACATCAGATGTTCCTCTCAGAAATCAAAACAAAGAGTTTCAAGAAAT 608
DB 846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsn 865
QY 609 TCTTGGGATTTCTGAGAGTCTCGGTGAGACTGTTTCAGAAGGATGTGTGTATCCCAAG 668
DB 866 SerTyrAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
QY 669 GCTTACACATCAAAAGAAATGGAATAAAATAGTGGAAATTAAGAGATTCAACTAGCCCTA 728
DB 886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 905
QY 729 TCAAAATCTTGATACAGTTCAATCTTGTGAAAGCAAGGAACTTCAAAAAGATCAC 788
DB 906 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 925
QY 789 TGTGAACAACGTACAGGAAATGGAACAAATGAAAGAAAGTCTTGTGTACTGAAAGAAAG 848
```

```
DB 926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys 945
QY 849 AAAGTGTGAGAGCAGAAAGAAATAAATAACAGTTCAGAGTTCAGAGAACCAAAAGTTAAATGGGAA 908
DB 946 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGlnAsnGlnLysValLysTyrGlu 965
QY 909 CAAGAGCTCTGAGTGTGAGATTGACATTTTAAACCAAGAGAGAGAGAGAGAAATGCC 968
DB 966 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluGluLysArgArgAsnAla 985
QY 969 GATATATTAATGAANAATTAGGAGAAATTAGGAGAAATTAGGAGAAATGAGAGCAGTAGGAA 1028
DB 986 AspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGluGlnHisArgLys 1005
QY 1029 GAGTTAGAGTGAACCAACCAACTTGAACAGGCTCTCAGATACCAAGATATAGAAATGAAG 1088
DB 1006 GluLeuGluValLysGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 1025
QY 1089 AGTGTGAAAGTAAATTTGAATCAGGTTTCTCACATCATGAAATGAAATATATCTCTTA 1148
DB 1026 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 1045
QY 1149 CATGAAATTTGCATGTTGAAAAAGGAAATTCCTCATCTGCTAAACTGGAATAGCACACTG 1208
DB 1046 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 1065
QY 1209 AAACACCAATACCAAGGAAAGAAATAATACTTTGAGACATTAAGATTTTAAAGAA 1268
DB 1066 LysHisGlnTyrGlnGluLysGluAsnLysTy-PheGluAspIleLysIleLeuLysGlu 1085
QY 1269 AAGATCTCAACTCAGATGACCTAAACTCAAGAGAGGAATCATTAACTAAAGGCA 1328
DB 1086 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluSerLeuThrLysArgAla 1105
QY 1329 TCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGAACCAATGCTCACTTAA 1388
DB 1106 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1125
QY 1389 TTGAGGAAAAACAGCAAGAAATACTAGAGGCGAATTCGAATCACACCATCCTAGA 1448
DB 1126 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisLysProArg 1145
QY 1449 CTGCTTCTCTGTACAAAGACCATGATCAAAATGTGACATCAAGAAAAAGTCAAGAACCT 1508
DB 1146 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1165
QY 1509 GCTTTCACATTCGAGAGATGCTTTTCCAAAGAAAAATGAATGTTGATGAGTAGT 1568
DB 1166 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1185
QY 1569 ACCGATATATAACAATGAGGTGCTCCATCAACACACTTCTGAAAGCTCAAGGAAATCCAN 1628
DB 1186 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 1205
QY 1629 AAGCTTAAATTAATCTCAATTAATGAGAGATGCTCTTAAGAGAAAAATACATGTTGTTTC 1688
DB 1205 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 1225
QY 1689 AGGAACATGCACAAAGAGACCAACGTAACACAGTGTCAAAATGAAGGAGAGCTGGAACACA 1748
DB 1225 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHis 1245
QY 1749 TGTNTCAAAACGACAGATNATGTGAACAAAACACTGANCAGCAGGAGTCTCTAGATC 1808
DB 1245 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1265
QY 1809 AGAATTTATTCACATCAAGCAAAAATATGCGTTCACACAGCAATAGTTCATGCAC 1868
DB 1265 LysLeuPheGlnLeuGlnSerLysAsnMetTyrLeuGlnGlnLeuValHisAlaHis 1285
QY 1869 ATAANGAAAGCTGACAAACAAAGCAAGATAAACAATTCATNTTCATTTCTTGAGAGAAA 1928
```

```
Db 1285 is-LysLysAlaAspAsnLysSerLysLeuThrIleAspIleHisPheLeuGluArgLys 1304
QY 1929 ARGON-CATCATCTCTAAAGAGAAAATAGAGAGATATTNATTACATTAACCATTTA 1987
Db 1305 MetGlnHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1324
QY 1988 AAAACCCGTATATTCAATATGGAAGAAAAGAAAAGAAA 2027
Db 1325 LysAsn-ArgIleTyrGlnTyr-GluLysGluLysAlaGlu 1337

RESULT 2
QNSI9
ID Q9NSI9 PRELIMINARY; PRT; 424 AA.
AC Q9NSI9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PRE4 protein (Fragment).
GN PRE4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.S.,
RA Toyoda A., Ishii K., Totoki Y., Choi D.K., Soeda E., Ohki M.,
RA Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald A., Kump A., Schillabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordieck G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Vaspo M.L.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AL163203; CAB90394.1; -.
FT NON_TER 1
FT NON_TER 424
SQ SEQUENCE 424 AA; 50038 MW; C6E704795C534625 CRC64;

Alignment Scores:
Pred. No.: Length: 2,53e-87 424
Score: 1559.50 Matches: 334
Percent Similarity: 82.84% Conservative: 33
Best Local Similarity: 75.40% Mismatches: 55
Query Match: 45.83% Indels: 21
DB: Gaps: 3

US-09-602-362E-15 (1-2030) x Q9NSI9 (1-424)
QY 714 GATTCACTAGCTATCAAAAATCTGGATACAGTTCATCTTGTGAAGAGCAAGGAA 773
Db 2 AspSerThrLeuSerLysLeuAspAlaValProSerCysGluArgGlyLysGlu 21
QY 774 CTTCAAAAAGATCATCTGCAACACGTACAGGAAAATGGAACAAATGAAAAGAGTTT 833
Db 22 LeuLysLysAspHisCysGluGlnIleThrAlaLysMetGluGlnThrLysAsnLysPhe 41
QY 834 TGTGTACTGAAAAGAACTGTGAGAGCAAGCAAGAAAATAAATACAGTTAGAGAACCAA 893
Db 42 CysValLeuGlnLysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGln 61
QY 894 AAAGTTAAATGGAAACAAGAGCTCTGCAGTGTGAGATTGACTTTAAACCAAGAAAGAG 953
Db 62 LysAlaLysTrpGluGlnGluCysSer-----Lys 72
QY 954 AAGAGAGAAATGCCCATATATTAATGAAAATTAGGAAGAAATTAGGAAGATCGAA 1013
Db 73 LysArgArgAspValAspIleLeuLysGluLysIleArgPro-----Glu 87
```

```
QY 1014 GAGCGCATAGGAAAGAGTTAGAGTGAAACAACTTGAACAGGCTCTCAGAAATACAA 1073
Db 88 GluGlnLeuArgLysLysLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgLysGln 107
QY 1074 GATATAGAATTGAAGAGTGTAGAAAGTAATTTCAATCAGGTTTCTCAGCTCATGAAT 1133
Db 108 AspIleGluLeuLysSerValThrSerAsnLeuAsnGlnValSerHisThrHisGluSer 127
QY 1134 GAAAATTATCTCTACATGAAATTCATGTTTCAAAAGGAAATTCCTCATGCTTAAACTG 1193
Db 128 GluAsnAspLeuPheHisGluAsnCysMetLeuLysGluIleAlaMetLeuLysLeu 147
QY 1194 GAAATAGCCACACTGAAACCAATACAGGAAAAGGAAAATAAATACTTTTGAGGCATT 1253
Db 148 GluValAlaThrLeuLysArgGlnHisGlnValLysGluAsnLysTyrPheGluAspIle 167
QY 1254 AAGATTTTAAAGAAAAGAAATGCTGAATTCAGATGACCCCTAAACTGAAGAGGAATCA 1313
Db 168 LysIleLeuGlnGluLysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGlnLysThr 187
QY 1314 TTAACATAAAGGGCATCTCAATATATAGTGGCAGCTTAAAGTTTCTGATAGCTGAGAACAC 1373
Db 188 LeuThrLysArgAlaSerGlnTyrArgGluGlnLeuLysValLeuThrAlaGluAsnThr 207
QY 1374 ATGCTCACTTCTAAATTGAAGGAAAACAAAGACAAAGAAATATCTAGAGCGAGAAATGAA 1433
Db 208 MetLeuThrSerLysLeuLysGluLysGlnAspLysGluIleLeuGluThrGluIleGlu 227
QY 1434 TCACACCATCTAGATGGCTTCTGCTGTACAAGACCATCATCAAAATTTGTGACATCAAGA 1493
Db 228 SerHisHisProArgLeuAlaSerAlaLeuGlnAspHisAspGlnSerValThrSerArg 247
QY 1494 AAAAGTCAAGAACCTGTCTTCCCATTCGAGGAGATCTTGTGTTGCAAGAAAATAATGAAT 1553
Db 248 LysAsnGlnGluLeuAlaPheHisSerAlaGlyAspAlaHisLeuGlnGlyIleMetAsp 267
QY 1554 GTTCATGTGATGATACCGATATATACATGAGGTCTCTCATCAACCATCTTCTGAGG 1613
Db 268 ValAspValSerAsnThr-IleTyrAsnAsnGluValLeuHisGlnProLeuTyrGluAl 287
QY 1614 TCAAGGAAAATCCANAAGCCCTAAAAATTAATCTCAATTATGCGAGGAGATCTCTTAAGAGA 1673
Db 287 aGlnArgLysSerLysSerProLysIleAsnLeuAsnTyrAlaGlyAspAspLeuArgG 307
QY 1674 AAATACATCTGTTTCAGGACATGCAACAAAGAGACCAACGCTGAAACACACAGTGTCAATGA 1733
Db 307 uAsnAlaLeuValSer-GluHisAlaGlnArgAspArgCysGluThrGlnCysGlnMetL 327
QY 1734 AGGAAGTGAACACATCTNTCAAACGAAACAAGATNATGTGAACAAACACACTGANCAGC 1793
Db 327 yLysAlaGluHisMetTyrGlnAsnGluGlnAspAsnValAspLysHisThrGluGlnG 347
QY 1794 AGGAGTCTTAGATCAGAAATATTTCAACTACAAGCAAAAATATGTGCGCTTCACAGC 1853
Db 347 InGluSerLeuGluGlnLysLysPheLysLeuGluSerLysAsnArgTrpLeuArgGlnG 367
QY 1854 AATTAGTTCATGACATAANGAAGCTGACAAACAAAGCAAGATAACAATTGATTTTCAT 1913
Db 367 InLeuValTyrAlaHis-LysLysVal---AsnLysSerLysValThrIleAsnLysGln 385
QY 1914 TMTCTTGAGAGGAAAATGCNCATCATCTTCTAAAGAGAAAATGAGGAGATATTNATT 1973
Db 386 PheProGluThrLysMet--GlnArgHisLeuLysGluLysAsnGluGluValPheAsnT 405
QY 1974 ACNATACCATTTTAAAAACCCGTTATTTCAATATGGAAGAAAAGAAAAGAAAAGAAA 2030
Db 405 yrGlyAsnHisLeuLysGlu-CysIleAspGlnTyrGluLysGluLysAlaGluArg 423

RESULT 3
Q9BXX2
ID Q9BXX2 PRELIMINARY; PRT; 1011 AA.
AC Q9BXX2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
```



```
QY 1115 ----- 1115
Db 859 AspAlaGlnArgInLeuSerArgGluGlnAsnAlaArgMetLeuGlnAspGlyIleLeu 878
QY 1115 ----- 1115
Db 879 ThrAsnHisLeuSerLysGlnLysGluIleGluMetAlaGlnLysLysMetAsnSerGlu 898
QY 1116 ---TCTCACACTCATGAATGAAATATCTCTTACATGAAATGCGATGTTGAAAG 1172
Db 899 AsnSerHisSerHisGluGluGluLysAspLeuSerHisLysAsnSerMetLeuGlnGlu 918
QY 1173 GAAATGCGCATGCTAAACTGGAATAGCCACACTGAAACACCAATPACCAGAAAGGAA 1232
Db 919 GluIleAlaMetLeuArgLeuGluIleAspThrIleLysAsnGlnAsnGlnLysGlu 938
QY 1233 AATAAATCTTGGAGACATTAAGATTTAAAGAAAGAAAGATGCTGCACTTCAGATGACC 1292
Db 939 LysLysCysPheGluAspLeuLysIleValLysGluLysAsnGluAspLeuGlnLysThr 958
QY 1293 CTAATACTGAAAGAGGAATCATTAATAAGGAGGATCTCAATATAGTGGCAGCTTAA 1352
Db 959 IleLysGlnAsnGluGlnThrLeuThrGlnThrIleSerGlnTyraSngLysArgLeuSer 978
QY 1353 GTTCTGATAGTGAACACATGCTCACTTCTTAATTTGAAG---GAAAAACAGACAAA 1409
Db 979 ValLeuThrAlaGluAsnAlaMetLeuAsnSerLysLeuGluAsnGlnLysSerLys 998
QY 1410 GAAATCTAGAGGACAGAAATTTGAATCACACCATCTCTGCTCTCTCTCAAGAC 1469
Db 999 GluArgLeuGluAlaGluValGluSerTyHisSerArgLeuAlaAlaIleHisasp 1018
QY 1470 CATGATCAATTTGTGCATCAAGAAACAACTCAAGAACCTGCTTCCATTCACATTCAGAGAT 1529
Db 1019 ArgAspGlnSerGluThrSerLysArgGluLeuLeuAlaPheGlnArgAlaArgasp 1038
QY 1530 GCTTGT-----TTGCAAGAAAGAAATGAATTTGATGTGAGTAGTACCGATATACAA 1583
Db 1039 GluCysSerArgLeuGlnAspLysMetAsnPheAspValSerAsnLeuLys-AspAsnAs 1058
QY 1584 TGAGGTGCTCCATCAACCACTTTCTGAAGCTCAAAGGAAATCCANAAGCCTTAAATTTAA 1643
Db 1058 nGluIleLeuSerGlnGlnLysPheLysThrGluSerLysLeuAsnSerLeuGluIleG 1078
QY 1644 TCTCAATTTATGAGGAGAGTCTTAAGAGAAATACATTTGTTTCCAGAACATGCACAAA 1703
Db 1078 uPheHisHisThrArgAspAlaLeuArgGluLysThrLeu-GlyLeuGluArgValGlnL 1098
QY 1704 GAGACCAACGTGAAACACAGTGTCAAATGAAGGAGCTGCAACATGTTCAAACGCAAC 1763
Db 1098 ysAspLeuSerGlnThrGlnCysGlnMetLysGluMetGluGlnLysTyrglnAsnGluG 1118
QY 1764 AGATNATGTGAACAAACACACTGANCACGAGTCTCTAGATCAGAAATTTATTTCAAC 1823
Db 1118 InValLysValAsnLysTyriIleGlyLysGlnGluSerValGluGluArgLeuSerGlnL 1138
QY 1824 TCAAAAGCAAAATATGTGCTTCAACAGCAATTTAGTTTCATGCACATAANGAAGCTCAC 1883
Db 1138 euGlnSerGluAsnMetLeuLeuArgGlnGlnLeuAspAlaHis-AsnLysAlaAsp 1157
QY 1884 ACAAAAGCAAGATAACAATTGAT-----NTTCATNTTCTTGAGAGGAAATG 1931
Db 1158 AsnLysGluLysThrValIleAsnIleGlnAspGlnPheHisAlaIleValGlnLysLeu 1177
QY 1932 CNCATCAT-----CTTCAAGAGAGAAATGAGGAGATATTTTAT 1972
Db 1178 GlnAlaGluSerGluLysGlnSerLeuLeuLeuLeuGluArgAsnLysGluLeuLeuSer 1197
QY 1973 TACNATAACATTTAAAAACCCGTATATTTCAATATGGAAGAAAAAANAAAAA 2030
Db 1198 GluCysAsnHisLeuLysGlu-ArgGlnTyrglnTyrgluAsnGluLysAlaGluArg 1216
```

```
RESULT 5
Q9H1Q1
ID Q9H1Q1 PRELIMINARY; PRT; 1710 AA.
AC Q9H1Q1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE BAI4588.1 (Hypothetical protein KIAA1074).
GN BAI4588.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bird C.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162272; CAC19649.1; -.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 4.
DR SMART; SM00248; ANK; 4.
DR PROSITE; PS00086; ANK_REPEAT; 4.
DR PROSITE; PS0297; ANK_REPEAT_REGION; 1.
KW Hypothetical protein; ANK repeat; Repeat.
SQ SEQUENCE 1710 AA, 196410 MW, 01CBF9BADB894872 CRC64;

Alignment Scores:
Pred. No.: 1e-64 Length: 1710
Score: 1188.00 Matches: 309
Percent Similarity: 56.62% Conservative: 123
Best Local Similarity: 40.50% Mismatches: 213
Query Match: 34.91% Indels: 119
DB: 4 Gaps: 16
```

```
US-09-602-362E-15 (1-2030) x Q9H1Q1 (1-1710)
QY 33 AACTGCGGAATGAAAGTTTCTATTCCAACTAAAGCCTTAGAATTGTGACATGCAAACT 92
Db 479 AsnValGlyMetProValAla----- 485
QY 93 TTCAAGACAGAGCTCCGAGAAAGCCATTCGCTTCGAGCTGCCATTGAAATGCAAAAG 152
Db 486 ---HisMetGluSerProGluArgTyLeuHisLeuLysProThrIleGluMetLysasp 504
QY 153 TCTGTTCCAAATAAGCTTGAATTCAGAAATGAACAAACATTTGAGACAGATGAG--- 209
Db 505 SerValProAsnLysAlaGlyMetLysAspValGlnThrSerLysAlaAlaGluHis 524
QY 210 ATACTCCCATCAGAATCCAAACAAAGGACTATGAAGAAAGTTCTTTGGGATTCGAGAT 269
Db 525 AspLeuGluValAlaSerGluGluGlnGluArgGluGlySerGluAsnAsnGlnPro 544
QY 270 CTCTGTGAGACTTTTCACAGAGGATGTGTGTTTACCACAGGCTACACATCAAAAGAA 329
Db 545 GlnValGluGluArgLysLys-----HisArgAsnAsnGluMetGlu 559
QY 330 ATA-----GATAAATAAATGCAAAATTAGAAGAGTCTCCTCGATTAATGAT 374
Db 560 ValSerAlaAsnIleHisaspGlyAlaThrAspAspAlaGluAspAspAspAsp 579
QY 375 GGTTTTCGTG-----AAGGCTCCCTCGAGNATGAAAGTTTCTATTCCAACTAAA--- 422
Db 580 GlyLeuIleGlnLysArgLysSerGlyGluThrAspHisGlnGlnPheProArgLysGlu 599
QY 423 -----GCTTAGAATTGATGCACATGCAA---ACTTTC 452
Db 600 AsnLysGluTyraLaserSerGlyProAlaLeuGlnMetLysGluValLysSerThrGlu 619
QY 453 AAAGCAGAGCTCCCGAGAGCCATTCGCTTCGAGCCTGCCATTCGAAATGCAAAAGTCT 512
Db 620 LysGluLysArgThrSerLysGluSerValAsnSerProValPheGlyLysAlaSerLeu 639
QY 513 GTTCCAAATAAAGCCTTGAATTTGAAGATTAACAAACATTTG----- 554
```



```

Db 640 LeuThrGlyGluLeuGlnValAspAspSerSerLeuSerGluLeuAspGluAsp 659
QY 555 -----AGAGCAGATCAGATGTCCTTCAGAAATCAAAACAAAAGAGAGTTGAAGAAAAT 608
Db 660 GluGlyArgProThrLysLysThrSerAsnGluLysAsnLysValLysAsnGlnLeuGln 679
QY 609 TCTTGGGATTCT-----GAGAGTCTCCGTGAGACTGTTTCACAGAGGATGTG 656
Db 680 SerMetAspAspValAspAspLeuThrGlnSerSerGluThrAlaSerGlu---AspCys 698
QY 657 TGTGTACCAACAGGCTACATCAACAAAGAAATG-----GATAAAATAAGTCGAAA 707
Db 699 GluLeuProHisSerSerTyLysAsnPheMetLeuLeuLeuGluGlnLeuGlyMetGlu 718
QY 708 TTAGAAGATTCAACTAGGCTATCAAAATCTTGATGATCAGGTTCAATCTTGTTGAAGAGCA 767
Db 719 CysLysAspSerValSerLeuLysIleGlnAspAlaLeuSerCysGluArgLeu 738
QY 768 AGGGAACCTTCAAAAGATCACTGTGAACAACGTCAGGAAAAATGGACAAATCAAAAG 827
Db 739 LeuGluLeuLysLysAsnHisCysGluLeuLeuThrValLysIleLysLysMetGluAsp 758
QY 828 AAGTTTCTGTACTGAAAAGAACTGTCAAGACAAAGAAATAAAATCAATCAGTTAGAG 887
Db 759 LysValAsnValLeuGlnArgGluLeuSerGluThrLysGluLeuLysSerGluLeuGlu 778
QY 888 AACCAAAAGATTAAATGGGAACAGAGCTCGCAGTGTGAGATTCACTTTAAACCAAGAA 947
Db 779 HisGlnLysValGluTrpGluArgGluLeuCysSerLeuArgPheSerLeuAsnGlnGlu 798
QY 948 GAAGAGAAGAGAAATCCCGATATATTAAATGAAAAATATAGGAGAAATAGGAAGA 1007
Db 799 GluGluLysArgArgAsnAlaAspThrLeuTyrgluLysIleArgGluGlnLeuArgArg 818
QY 1008 ATCGAAGACGACATAGAAAGAGTTAGAGTGAACCAACACTTGAACGGCTCTCAGA 1067
Db 819 LysGluGluGlnTyrgluValGluValLysGlnGlnLeuGluLeuSerLeuGln 838
QY 1068 ATACAAGATTAGAAATTCAGAGCTGAGAAAGTAAATTTGAATCAGGTT----- 1115
Db 839 ThrLeuGluMetGluLeuArgThrValLysSerAsnLeuAsnGlnValValGlnGluArg 858
QY 1115 ----- 1115
Db 859 AsnAspAlaGlnArgGlnLeuSerArgGluGlnAsnAlaArgMetLeuGlnAspGlyIle 878
QY 1115 ----- 1115
Db 879 LeuThrAsnHisLeuSerLysGlnLysGluLeuGluMetAlaGlnLysLysMetAsnSer 898
QY 1116 -----TCTCACACTCATCAAAATGAAATATTCTTTACATGAAATTCATGTTGAAA 1169
Db 899 GluAsnSerHisSerHisGluGluLysAspLeuSerHisLysAsnSerMetLeuGln 918
QY 1170 AAGGAATTCATGCTCAAACTGAAATAGCCACTGAACACCAATACAGGAAAG 1229
Db 919 GluGluIleAlaMetLeuArgLeuGluIleAspThrIleLysAsnGlnAsnGlnGluLys 938
QY 1230 GAAATAATAATCTTTGAGACATTAAAGATTTTAAAGAAAGAAATGCTGCAACTCAGATG 1289
Db 939 GluLysLysCysPheGluAspLeuLysIleValLysGluLysAsnGluAspLeuGlnLys 958
QY 1290 ACCCTAAACTGAAAGAGGAATCATTAATAAGGGCATCTCAATATAGTGGCAGCTT 1349
Db 959 ThrIleLysGlnAsnGluGluThrLeuThrGlnThrIleSerGlnTyrgluAsnGlyArgLeu 978
QY 1350 AAGATTCTGATAGCTGAGAACCAATGCTCACTTCTTAAATTTGAAG---GAAAAACAAGAC 1406
Db 979 SerValLeuThrAlaGluAsnAlaMetLeuAsnSerLysLeuGluAsnGluLysGlnSer 998
QY 1407 AAGAAATATACTAGAGCGAAATTAATCAATCACCATTCTAGACTGCTTCTGCTGTACAA 1466

```

```

Db 999 LysGluArgLeuGluAlaGluValGluSerTyHisSerArgLeuAlaAlaIleHis 1018
QY 1467 GACCATGATCAAAATTTGTGACATCAAGAAAAGTCAAGAACCTGCTTTCACATTCGAGA 1526
Db 1019 AspArgAspGlnSerGlnThrSerLysArgGluLeuGluLeuAlaPheGlnArgAlaArg 1038
QY 1527 GATGCTTGT-----TTGCAAAAGAAAATGAATGTTGATGTGATGATACCGCATATATA 1580
Db 1039 AspGluCysSerArgLeuGlnAspLysMetAsnPheAspValSerAsnLeuLys-AspAs 1058
QY 1581 CAATGAGTGTCTCATCAACCACTTTCTTGAAGCTCAAGAGAAATCCANAAGCTTAAAT 1640
Db 1058 nAsnGluIleLeuSerGlnGlnPheLysThrGluSerLysLeuAsnSerLeuGluI 1078
QY 1641 TAATCTCAATTATGACGAGATGCTCTTAAGAGAAAATACATGTTTCAGGAACATGCAC 1700
Db 1078 eGluPheHisHisThrArgAspAlaLeuArgGluLysThrLeu-GlyLeuGluArgValG 1098
QY 1701 AAAGAGACCAACCGTGAACACACAGTGTCAAAATGAAGAACTGAACACATGTGTTCAAACNG 1760
Db 1098 InLysAspLeuSerGlnThrGlnCysGlnMetLysGluMetGluGlnLysTyrglnAsnG 1118
QY 1761 AACAAAGATGATGAAACACACACATGANCACAGGAGTCTCTAGATCAGAAATATTTC 1820
Db 1118 InGlnValLysValAsnLysTyrglyLysGlnGlnSerValGluGluArgLysSerG 1138
QY 1821 AACTACAAAGCAAAATATGTGCTTCAACAGCAATAGTTCATGTCACATAANGAAAGCT 1880
Db 1138 InLeuGlnSerGluAsnMetLeuLeuArgGlnGlnLeuAspAspAlaHis-AsnLysAla 1157
QY 1881 GACACAAAGCAAGATAACATTGAT-----NTTCATTNTCTTGAGAGAGAA 1928
Db 1158 AspAsnLysGluLysThrValIleAsnIleGlnAspGlnPheHisAlaIleValGlnLys 1177
QY 1929 ATGCNCAATCAT-----CTTCTAAAGAGAAAAATCAGGAGATATTT 1969
Db 1178 LeuGlnAlaGluSerGluLysGlnSerLeuLeuLeuGluGluArgAsnLysGluLeuIle 1197
QY 1970 NATTCNATTAACCATTTAAACCCGCTATATTTCAATATGGAAGAAAAAANAANA 2029
Db 1198 SerGluCysAsnHisLeuLysGlu-ArgGlnTyrglnTyrgluAsnGluLysAlaGluAr 1217
QY 2030 A 2030
Db 1217 g 1217
RESULT 6
Q9HCD1 PRELIMINARY; PRT; 718 AA.
ID Q9HCD1 Q9HCD1
AC Q9HCD1; 2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, last annotation update)
DE Hypothetical protein KIAA1641 (Fragment).
GN KIAA1641.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20450683; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirotsawa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVIII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.";
RL DNA Res. 7:273-281(2000).
DR EMBL; AB046861; BAB13467.1; -.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 718 AA; 82972 MW; 2AD05ADE70E3C97F CRC64;
Alignment Scores:

```

| | | | |
|------------------------|----------|---------------|----|
| Pred. No.: | 1.83e-40 | Length: | 71 |
| Score: | 790.50 | Matches: | 23 |
| Percent Similarity: | 53.83% | Conservative: | 13 |
| Best Local Similarity: | 34.37% | Mismatches: | 24 |
| Best Local Similarity: | 34.37% | Indels: | 71 |
| Query Match: | 23.23% | Gaps: | 16 |
| DB: | 4 | | |

US-09-602-362E-15 (1-2030) x 09HCD1 (1-718)

[illegible]

| | | | |
|----|------|---|------|
| QY | 1011 | GAAGAGCAGCATAGAAAGAGTGTAGAGTGAACACACAACTTGAACAGGCTCTCAGNATA | 1071 |
| DB | 316 | GlulGlnGlnTyrArgIleGluAlaAspValThrLysProIleLysProAlaLeuLysSer | 335 |
| QY | 1071 | CAAGATATAGAATTGAAGAGCTGTAGAAAGTAATTGAAATCAGGTCTTCTCACATCTCATGAA | 1130 |
| DB | 336 | AlaGluValGluLeuLysThrGlyGlyAsnAsnSerAsnGlnValSerGluThrAspGlu | 355 |
| QY | 1131 | AATGAAATATCTCTTACATGAAATTCATGCTTGAAGAAAGAAATGCGCATGCTCAAAA | 1190 |
| DB | 356 | LysGluAsp---LeuLeuHisGluAsnArgLysMetGlnAspGluIleAlaArgLeuArg | 374 |
| QY | 1191 | CTGGAATAGCCACACTGAAACACCACCAATACCAGGAAAGGAAATAATACTTTGAGGAC | 1250 |
| DB | 375 | LeuGluLysAspThrIleLysAsnGlnAsnLeuGluLys-----LysTyrLeuLysAsp | 392 |
| QY | 1251 | ATTAAAGATTTTAAAGAAAGAATGCTGAACCTTCAGATGCACCTTAAACTGAAAGAGGAA | 1310 |
| DB | 393 | PheGluIleValLysArgLysHisGluAspLeuGlnLysAlaLeuLysArgAsnGlyGlu | 412 |
| QY | 1311 | TCATTAACTAAAAGGCGCATCTCAATATAGTGGCAGCTTAAAGCTTCTGATAGCTGAGAAC | 1370 |
| DB | 413 | ThrLeuAlaLysThrIleAlaCysTyrSerGlyGlnLeuAlaAlaLeuThrAspGluAsn | 432 |
| QY | 1371 | ACAATGCTCACTTCTAAATG---AGGAAAAACAGACAAAGAAATACTAGAGGCGAGAA | 1427 |
| DB | 433 | ThrThrLeuArgSerLysLeuGluLysGlnArgGluSerArgGlnLeuThrGlu | 452 |
| QY | 1428 | ATTGAATCACACCATCCTAGACTGGCTTCTGCTGTACAGACCATCATCAAAATTGTGACA | 1487 |
| DB | 453 | MetGlnSerTyrHisCysArgLeuAsnAlaAlaArgCysAspHisAspGlnSerHisSer | 472 |
| QY | 1488 | TCAGAAAAAGTCAAGAACCTGCTTTCCCATTCAGAGGAGTCTTGT-----TTGCAA | 1541 |
| DB | 473 | SerLysArgGlnGlnLeuAlaPheGlnGlyThrValAspLysCysArgHisLeuGln | 492 |
| QY | 1542 | AGAAAAATGAATGTTGATGTAGTAGTACCGATATATACAATGAGGTGCTCATCAACC | 1601 |
| DB | 493 | GluAsnLeuAsnSerHisVal-Leu-----IleLeuSerLeuGlu | 505 |
| QY | 1602 | ACTTTCTGAGCTCAAGGAATCCANAAGCTTAAAAATTAATCTCAATTATGACGAGA | 1661 |
| DB | 505 | nLeuSerLysAlaGluSerLysSerArgValLeuLysThrGluLeuHisTyrThrGlyGlu | 525 |
| QY | 1662 | TGCTCTAAGAGAAAAATACATTGGTTTCAGGAACATGCACAAAGAGACCAACGTGAACAC | 1721 |
| DB | 525 | uAlaLeuLysGluLysAlaLeuValPhe-GluHisValGlnSerGluLeuLysGlnLysG | 545 |
| QY | 1722 | AGTGCTCAATGAAGGAGCTGAACACATGNTCCAAAGCAACAAGATNATGTCAACAAAC | 1781 |
| DB | 545 | InsSerGlnMetLysAspIleGluLysMetLysSerGlyTyrAsnThrMetGluLysC | 565 |
| QY | 1782 | ACACTGANCAGCAGGAGTCTCTAGATCAGAAATATTTCAACTACAAAGCAAAAATATGT | 1841 |
| DB | 565 | ysIleGluLysGlnGlu-----ArgPheCysGlnLeuLysLysGlnAsnMetL | 581 |
| QY | 1842 | GGCTTCAACAGCAATTAGTTTCATGCACATAAGAAAGCTGACACAAAGCAAGATAACA | 1901 |
| DB | 581 | eLeuGlnGlnGlnLeuAspAspAlaArg-AsnLysAlaAspAsnGlnGluLysAlaIle | 600 |
| QY | 1902 | ATTGATNTTCATNTT-----CTTGAG-----AGGAAA | 1928 |
| DB | 601 | LeuAsnIleGlnAlaArgCysAspAlaArgValGlnAsnLeuGlnAlaGluCysArgLys | 620 |
| QY | 1929 | ATGCNCAATCATCTTCTAAAGAGAAAAATCAGAGATATTTTATTACNATAACCATTTAA | 1988 |
| DB | 621 | HisArgLeu--LeuLeuGluGluAspAsnLysMetLeuValAsnGluLeuThrHisSerL | 640 |
| QY | 1989 | AAAAACCGTATATTTCAATATGGAATAAAAAAANAANAANAANA | 2030 |
| DB | 640 | ysGlu-LysGluCysGlnTyrGluLysGluLysAlaGluArg | 653 |

RESULT 7

| | | | |
|----------|--|--|--------------|
| Qy | 1944 | TAAAGAGAAAATGACGAGATATTTTATTCNATAACCATTTAAAAAACCCGTATATTT | 2003 |
| | | | |
| Db | 399 | euGlnGiulysAsnLysGluLeuMetAspGluTyrAsnHisLeuLysGlu-ArgMetAsp | 418 |
| | | | |
| Qy | 2004 | CAATATGGAATAAAAAAABA | 2021 |
| | | | |
| Db | 419 | GlnCysGluLysGluLys | 424 |
| | | | |
| RESULT 9 | | | |
| Q8IZM7 | | | |
| ID | Q8IZM7 | PRELIMINARY; | PRT; 453 AA. |
| AC | Q8IZM7; | | |
| DT | 01-MAR-2003 | (TrEMBLrel. 23, Created) | |
| DT | 01-MAR-2003 | (TrEMBLrel. 23, Last sequence update) | |
| DT | 01-MAR-2003 | (TrEMBLrel. 23, Last annotation update) | |
| DE | Melanoma-associated antigen (Fragment). | | |
| OS | Homo sapiens (Human). | | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| OX | NCBI_TaxID=9606; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RA | Bruno R., d'Orlando O., Altomonte A., Lama J. E., Maio M., Pucillo C.; | | |
| RT | "Study of the immune profile in metastatic melanoma patients immunized | | |
| RT | with anti-idiotypic antibody by SEREX analysis." | | |
| RL | Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases. | | |
| DR | EMBL; AF543495; AAK40505.1; -. | | |
| DR | [1] | | |
| FT | NON TER | | |
| SQ | SEQUENCE | 453 AA; 53140 MW; 81DF80012122BBCA | CRC64; |

```

DR Pfam: PF00023; ank; 6.
DR SMART: SM00248; ANK; 6.
DR PROSITE; PS00088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REF_REGION; 1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 996 AA; 116065 MW; FDS8060E3D8A1543 CRC64;

Alignment Scores:
Pred. No.: 2,53e-19 Length: 996
Score: 443.50 Matches: 131
Percent Similarity: 50.84% Conservative: 80
Best Local Similarity: 31.57% Mismatches: 164
Query Match: 13.03% Indels: 40
DB: 4 Gaps: 6

US-09-602-362E-15 (1-2030) x Q8IVF6 (1-996)
QY 815 ACAATGAAAGAAAGTGTGTGCTGTAAGAAAGAAAGTGTGAGA----- 859
Db 239 ThrAlaGluAspTyrAlaLeuCysSerAspLeuArgSerIleArgGlnGlnIleLeuGlu 258
QY 860 -----AGCAAGAAAGAAATAAATACAGATTAGAGAACCAAAAGTTAAATGCGAACAAAGA 913
Db 259 HisLysAsnLysMet-LeuLysAsnHisLeuArgAsnAsp----- 271
QY 914 GCTCTGCAGTGTGAGATTGACTTTAAACCAAGAAAGAGAGAGAGAAATGCCGATAT 973
Db 272 -----AsnGlnGluThrAlaAlaMetLysProAla----- 281
QY 974 ATTAAATGAAAGAAATAGGAAGCAATTAGGAAGATCGAAGAGACAGATAGAAAGAGTT 1033
Db 282 -----AsnLeuLysLysArgLysGluArgAlaLysAlaGluHisAsnLeuLysValAla 300
QY 1034 AGAAGTGAACCAACAACTGACAGCGCTCTCAGAAATACAAGATATAGAAATGGAAGTGT 1093
Db 300 rGluGluLysGlnGlu-----ArgLeuGlnArgSer 310
QY 1094 AGAAGTAATTGAATCAGGTTTCTCAGCTCATGAAATGAAATGAAATATCTCTTACATGA 1153
Db 310 rGluAsnLysGlnProGlnAspSerGlnSerTyrGlyLysLysLysAspAlaMetTyrG 330
QY 1154 AAATTGCTATTTGAAAGAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1213
Db 330 yAsnPheMetLeuLysLysAspIleAlaMetLeuLysGluGluLysLysAlaLysAs 350
QY 1214 CCAATACCGAGAAAGAAATTAATCTTTGAGGACATTAAGATTTTAAAGAAAGAAAGNA 1273
Db 350 nAspSerLeuArgLysGluLysLysTyrIleGlnGluIleLysSerIleThrGluLea 370
QY 1274 TGCTGAACTTCAGATGACCTCAAACTGAAAGAGGAATCATTAACCTAAAGGGCATCTCA 1333
Db 370 nAlaAsnPheGluLysSerValArgLeuAsnGluLysMetIleThrLysThrValAla 390
QY 1334 ATATAGTGGCGACCTTAAAGTTCTGATAGCTGAGAACACAACTGCTCACTTCTAAATG-- 1391
Db 390 gTyrSerGlnGlnLeuAsnAspLeuLysAlaGluAsnAlaArgLeuAsnSerGluLeu 410
QY 1392 -AAGGAAACCAAGCAAGAAATATCTAGAGGAGAGAAATGAAATCAACATCCATCTAGACT 1450
Db 410 uLysGluLysHisAsnLysGluArgLeuGluAlaGluValGluSerLeuHisSerLeu 430
QY 1451 GGCTTCTGCTGTCAAGACCATCATCAATTTGTGACATCAAGAAAGTCAAGAACCTGCTG 1510
Db 430 uAlaThrAlaIleAsnGluLysAsnGluIleVal--GluArgLysAspLeuGluLeuVa 449
QY 1511 TTCCCATTCGAGGAGATCTTGTTCGAAAGAAATGAATGTTGATGCTGAGTAGTAC 1570
Db 449 lLeuTrpArgAlaAspValSerArgHisGluLysMetGlySerAsnIleSerGlnLe 469
QY 1571 CGATATATAACAATGAGTGTCTCCATCAACCACTTTCTGAAGCTCAAGAAATCCANAA 1630
Db 469 u-ThrAspLysAsnGluLeuThrGlnGlnValHisLysAlaArgValLysPheAsnT 489

```

```

QY 1631 GCCTAAAAATTAATCTCAATTATCGAGGAGATCTCTTAAGAGAAAAATACATTCGTTTCAG 1690
Db 489 hrLeuLysGlyLysLeuArgGluThrArgAspAlaLeuArgGluLysThrLeuAlaLeuG 509
QY 1691 GAACATGCACAAAGAGACCACGCTGAACACACAGTGTCAATGAGGAGCTGAACACATG 1750
Db 509 lySer-ValGlnLeuAspLeuArgGlnAlaGlnHisArgIleLysGluMetLysGlnMet 528
QY 1751 TTTCAAAACGAAACAAAGATNATGTCAACAAACACACATGACGAGGAGTCTCTAGATCAG 1810
Db 529 HisProAsnGlyGluAlaLysGluSerGlnSerIleGlyLysGlnAsnSerLeuGluGlu 548
QY 1811 AATATTATTCACATCAACAAACAAATATGCTTCAACAGCAATAGTTTCATGTCACAT 1870
Db 549 ArgIleArgGlnGlnGluLeuGluAsnLeuLeuGluArgGlnLeuGluAspAlaArg 568
QY 1871 AANGAAGCTGACAAACAAACAAAGATTAACAATGATNTTCATTTCTTTCAGAGGAAAAAT 1930
Db 569 LysGluGlyAspAsnLysGluIleValIleAsnIleHisArgAspCysLeuGluAsnGly 588
QY 1931 GCNCATCATCTTCTAAAGAGAAATAGGAGATATTTNATACNATACCATTTAAATA 1990
Db 589 LysGluAspLeuLeuGluGluArgAsnLysGluLeuMetLysGluTyrAsnTyrLeuLys 608
QY 1991 AACCCGTATATTCAATGCAAAAAAAGAAAAAANAANAANA 2029
Db 609 GluLysLeuLeuGlnCysGluLysGluLysAlaGluArg 621

RESULT 11
Q8T5C7 PRELIMINARY; PRT; 2055 AA.
AC Q8T5C7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Chimeric erythrocyte-binding protein MAEBL.
GN MAEBL.
OS Plasmodium falci-parum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Blair B.L., Kappe S.H.T., Maciel J.B., Balu D.J., Adams B.;
RT "Plasmodium falci-parum mael is a unique member of the ebl family.";
RL Mol. Biochem. Parasitol. 0:0-0(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Michon P., Stevens J.R., Kaneko O., Adams J.H.;
RT "Plasmodium falci-parum mael is a unique member of the EBL family.";
RL Mol. Biol. Evol. 0:0-0(2002).
RN [3]
RP SEQUENCE FROM N.A.
RA Ghai M., Dutta S., Ockenhouse C.F.;
RT "Identification, expression and functional characterization of MAEBL,
a sporozoite and asexual blood stage chimeric erythrocyte binding
protein of Plasmodium falci-parum.";
RL Mol. Biochem. Parasitol. 0:0-0(2002).
DR EMBL; AF042084; AAL10509.1; -.
DR EMBL; AF400002; AAM90625.1; -.
DR HSP; P04268; IIC2.
DR InterPro; IPR008602; Duffy binding.
DR Pfam; PF05424; Duffy_binding; 1.
SQ SEQUENCE 2055 AA; 243210 MW; 783F48BB7286B1F3 CRC64;

Alignment Scores:
Pred. No.: 1,98e-12 Length: 2055
Score: 330.00 Matches: 171
Percent Similarity: 42.69% Conservative: 124
Best Local Similarity: 24.75% Mismatches: 260
Query Match: 9.70% Indels: 136
DB: 5 Gaps: 27

```

US-09-602-362E-15 (1-2030) x Q8T5C7 (1-2055)

US-09-602-362E-15 (1-2030) X Q8T5C7 (1-2055)

| | | | |
|----|------|---|------|
| Qy | 15 | GATGGTCTTCTGAAGGCTAACTCGGAATGAAAGTTTCTATTCCAATAAGCCTTAGAA | 74 |
| Db | 1264 | AspGluLeuLysLysAlaGluGluLysLysAlaAspGluLeuLysLysSerGluGlu | 1283 |
| Qy | 75 | TTGATGGACATGCAAACTTTCAAGCAGAGCCTCCGAGAGCCATCTGCCTTCGAGCCT | 134 |
| Db | 1284 | LysLysLysAlaAspGluLeuLysLysAlaGluGluLysLysLysAlaAspGlu--- | 1302 |
| Qy | 135 | GCATTGAAATGCAAAAGTCTGTTCAAAATAAGACCTTGGAAATGAAGAT----- | 188 |
| Db | 1303 | --LeuLysLysLysAlaGluGluLysLysAlaAspGluLeuLysLysAlaGlu | 1321 |
| Qy | 189 | CAAACTTGAGACGAGTGAATCTCCATCAGATCCAAACAAAGACTATGAGAA | 248 |
| Db | 1322 | GluLysLysLysAlaAspGluValLysLysAlaGluGluLysLysLysAlaAspGluLeu | 1341 |
| Qy | 249 | AGTTCTTGGAGATCTCTGAGACTCTTCAGACATGTTTCACAGAAGGATGTGTGTTTACC | 308 |
| Db | 1342 | LysLysSerGluGluLysLysLysAlaAspGlu-----LeuLys | 1354 |
| Qy | 309 | AAGCTACACATCAAAAAGAAATAGATAAAATAATGGAAATTAGAAGAGTCTCCTGAT | 368 |
| Db | 1355 | LysSerGluGluLysLysLysAlaAspGluLeuLysLysLysAlaGluLysLysLys | 1374 |
| Qy | 369 | AATGATGTTTCTGAAGGCTCCTGAGAGTCTTCAGAACTTTCTATTCCAACTAAGCCTTA | 428 |
| Db | 1375 | AlaAspGluLeuLysLys-----LysAlaGluGluLysLysLysAlaAsp | 1389 |
| Qy | 429 | GAATTGATGACATGCAAACTTTCAAGCAGAGCCTCCGAGAGCCATCTGCCTTCGAG | 488 |
| Db | 1390 | GluLeuLysLys-----LysAlaGluGluLysLysLysAlaAspGluLeuLys | 1405 |
| Qy | 489 | CCTGCCATTGAAATGCAAAAGTCTGTCCAAAATAAGCCCTTGAATTTGAAGAT----- | 542 |
| Db | 1406 | LysLysAlaGluGluLysLys-----LysAlaAspGluLeuLysLysLysAla | 1421 |
| Qy | 543 | GAACAACATTGAGAGCAGATCAGATGTTCCCTTCAGAAATCAAAACAAGAGAGTTGAA | 602 |
| Db | 1422 | GluGluLysLysLysAlaAspGluLeu-----LysLysLysLysAlaGlu | 1435 |
| Qy | 603 | GAATAATCTTGGGATCTCGAGAGTCTCCGTGAGACTGTTTCACAGAAG----- | 656 |
| Db | 1436 | GluLysLysLys-----LysAlaGluAsnLeuLysLysAlaGluGluLysLysAlaAspGlu | 1454 |
| Qy | 657 | TGCTGATCCCAAGGCTACACATCAAAAAGAAATGATATAAATAAGTGAAATTAGAAGAT | 716 |
| Db | 1455 | LeuLysLysLysAlaGluGluLysLysAlaAspGluLeuLysLysLysAlaGluGlu | 1474 |
| Qy | 717 | TCAACTAGCCTATCAAAAATCTTGGATACAGTTTCATTCTTGTGAAGAGCAAGGAACCT | 776 |
| Db | 1475 | LysLysLysAlaAspGluLeuLysLysLysAlaGluGluLysLysLysAlaAspGluLeu | 1494 |
| Qy | 777 | CAAAAAGATCACTGTGAACAACGTCAGGAAAAATGGAA-----CAATGAAAAAGAG | 830 |
| Db | 1495 | LysLysAlaGluGluLysLysLysAlaAspGluLeuLysLysAlaGluGluLysLysLys | 1514 |
| Qy | 831 | TTTTGTGTACTGAAAAAG-----AACTGTCTCAGAGCAAGCAAAA | 866 |
| Db | 1515 | AlaAspGluLeuLysLysLysAlaGluGluLysLysLysAlaAspGluLeuLysLysAlaGlu | 1534 |
| Qy | 867 | GAAATAAATCAGCTTAGAAGAACCAAAAGATT-----AAATGGGAACAGAGCTC | 917 |
| Db | 1535 | GluLeuLysLysAlaGluGluLysLysValGluGlnLysLysAAGGluGluGluArg | 1554 |
| Qy | 918 | TGCAGTGTGAGATTG-----ACTTTAAACCAA----- | 944 |
| Db | 1555 | ArgAsnMetalLeuArgAlaGluLeuLysGlnLeuLysLysAlaGluLysLysArgIleGlu | 1574 |
| Qy | 945 | -----GAAGAAGAGAGAGAAATCCCGATATATTAAATGAAAAA | 986 |
| Db | 1575 | GluValMetLysLeuTyGluGluGluLysLysMetLysAlaGluGlnLeuLysLysGlu | 1594 |

| | | | |
|----|------|--|------|
| QY | 967 | ATTAGGGAAGAAATTAGGAAGAAATCGAAGACACATAGGAAGACAGTTAGAGTGAACAA | 1046 |
| DB | 1595 | Gluluglulysilelys-----Alagluglulneulysylsglugluglulyslys | 1612 |
| QY | 1047 | CAACTTGAACAGGCTCTCAGAATATACAGATATAGAATTGAAGAGTGTAGAACTAATTG | 1106 |
| DB | 1613 | LysValGluGlnLeulysLysLysGluGluGluGluLysLysLysAlaGluGlnLeulys | 1632 |
| QY | 1107 | AATCAGGTTTTCACACTCATGAATGAAATTAATCTCTTACATGAAAAATTCATGCTTG | 1166 |
| DB | 1633 | LysGlu-----GluGluGluAsnLysIleLysAlaGluGlnLeulysLys | 1647 |
| QY | 1167 | AAAAAGGAAATTCGCATCTAAACTGGAATAGCCACTGTAACACCACTAACCTACAGGAA | 1226 |
| DB | 1648 | LysGluGluGluLysLysLysAlaGlu-----GluLeuLysLysGluGlu-----Gln | 1664 |
| QY | 1227 | AAGGAAATAATACTTTTGAGGACATTAACATTTTAAAGAAAAAGAAATCTGAACCTTCAG | 1286 |
| DB | 1665 | GluluglulysLysAlaGluGlnLeulysLysGluGluGluGluLysLysValGlu | 1684 |
| QY | 1287 | ATGACCTTAAACTGGAAGGGAATCATTAACATAAAGGGCCTCTCAATATAGTGGCAG | 1346 |
| DB | 1685 | Gln---LeulysLysLysGluGluGluLysLysAlaGlu-----Gln | 1699 |
| QY | 1347 | CTTAAAGTCTGATAGCTGAGAACACATGCTCACTTCTAAATGGAAGAAAAACAAAGAC | 1406 |
| DB | 1700 | LeulysLysGluGluGluGluGluGluLysLysValGluGluGluLysLysGluGluGlu | 1719 |
| QY | 1407 | AAAGAAATCTAGAGGAGGAAATGGAATCACCATCTAGACTGGCTCTCTGCTGTACAA | 1466 |
| DB | 1720 | GluluglulysLysLysAlaGlu-----GluLeuLysLysGluGluGluGlu | 1733 |
| QY | 1467 | GACCATGATCAAAATGTGACATCAAGAAAAAGTCAAGAACTGCTTTCCATTGCAAGGA | 1526 |
| DB | 1734 | GlulysLysLysValGlnGlnLeulysLysGluGluGluLys----- | 1747 |
| QY | 1527 | GATGCTGTTTGAAAGAAAAAATGAATGTGATGTAGTAGTACCGATATATAACAATGA | 1586 |
| DB | 1747 | ----- | 1747 |
| QY | 1587 | GGTGTCCATCAACACTTTCTGAAGCTCAAAGGAAATCCANAAGCCTAAAAATTAATCT | 1646 |
| DB | 1748 | -----LysAlaGluGluIleArgLysGluLysGlu--Alav | 1759 |
| QY | 1647 | CAATTATGCAAGAGATGCTCTAAGAGAAATACATTTGGTTTCAGGAACATGCACAAAG | 1706 |
| DB | 1759 | allieGluGluGluLeu-----LysLysGluAspGluLysA | 1771 |
| QY | 1707 | ACCAACGTGAACACAGAGTCAAAATGAAGGAAGCTGAACACATGTCNAANCAACAAG | 1766 |
| DB | 1771 | rgArgMetGluValGluLysLysLysLysAspThrLysAspAsnPheGluAsnIleGlnG | 1791 |
| QY | 1767 | ATNATGTGAACAAACACT-----GANCAGAGGAGTCTTAGATCAGAAATATTTC | 1820 |
| DB | 1791 | luGlyAsnAsnLysAsnThrProTyrIleAsnLysGluMetPheAspSerGluIleLysG | 1811 |
| QY | 1821 | AACTACAA---AGCAAAAATATGTGGCTTCAACAGCAATTAAGTTCATGCACATANGAAA | 1877 |
| DB | 1811 | luValIleThrLysAsnMetGlnLeuAsnGluAlaAspAlaPheGluLysHisAsnS | 1831 |
| QY | 1878 | GCTGCAACAAAGCAAGATACAAATGATNTTCTATTCTTGTAGAGGAAATGCNCATC | 1937 |
| DB | 1831 | erGluAsnSerLysSerSerAsnLysAsnAlaAspPheSerLysGluLysAsp----- | 1848 |
| QY | 1938 | ATCTTTAAAAAGAGAAAAATCAGGAGATA | 1966 |
| DB | 1849 | --LeuLeuGluAspAspIleGluAsnIle | 1857 |

| |
|-----------|
| RESULT 12 |
| Q8IHP3 |
| ID Q8IH |
| AC Q8IH |

QY 1527 GATGCTTGTTCGAAAGAAAATGAATGTTGATGTGAGTAGTACCGATATATACAATGA 1586
 Db 1747 ----- 1747
 QY 1587 GGTGCTCCATCAACCACCTTCTGAGCTCAAGGAAATCCANAAGCCCTAAATAATTAATCT 1646
 Db 1748 -----LysalaGluGluIleAa-gLysGluLysGlu--Alav 1759
 QY 1647 CAATTATGAGGAGATGCTCTAAGAGAAAATACATTTGTTTCAGGAACATCCACAAGAG 1706
 Db 1759 alileGluGluLys-----LysLysGluAspGluLysA 1771
 QY 1707 ACCACGTGAACACAGTGTCAATGAAGGAGCTGAACACATGNTCAAACGACAAAG 1766
 Db 1771 tGArgMetGluValGluLysLysIleLysAspThrLysAspAsnPheGluAsnIleGlnG 1791
 QY 1767 ATNATGTGAACAAACACACT-----GANCACGAGGAGTCTCTAGATCAGAAATATTTC 1820
 Db 1791 luGluAsnAsnLysAsnThrProTyrIleAsnLysGluMetPheAspSerGluLysG 1811
 QY 1821 AACTCAAA---AGCAAAAATATGTGGCTTCAACAGCAATAGTTTCATGCACATANGAAA 1877
 Db 1811 luValIleThrLysAsnMetGlnLeuAsnGluAlaAspAlaPheGluLysHisAsnS 1831
 QY 1878 GGTGACAAACAAAGCAAGATACAATTGATNTTCATTCTTGAGAGGAAAATGCNCATC 1937
 Db 1831 erGluAsnSerLysSerSerAsnLysAsnAlaAspPheSerLysGluLysasp----- 1848
 QY 1938 ATCTCTTAAAGAGAAAATGAGGAGATA 1966
 Db 1849 --LeuLeuGluAspAspIleGluAsnIle 1857
 RESULT 13
 O76329 PRELIMINARY; PRT; 1738 AA.
 ID AC O76329; ID O76329; PRT; 1738 AA.
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Interactin.
 GN ABPD.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 ON NCBI_TaxID=44689;
 RN [1]_
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98365468; PubMed=9700162;
 RA Rivero F.J.; Kuspa A.; Brokamp R.; Matzner M.; Noegel A.A.;
 RT "Interactin, an actin-binding protein of the alpha-actinin superfamily
 RT in Dictyostelium discoideum, is developmentally and cAMP-regulated and
 RT associates with intracellular membrane compartments.";
 RL J. Cell Biol. 142:735-750(1998).
 DR EMBL; AF057019; AAC34582.1; --
 DR FIR; T14867; T14867.
 DR HSP; P46939; IQAG.
 DR GO; GO:0003779; F:actin binding; IEA.
 DR InterPro; IPR001589; Actbind actinin.
 DR InterPro; IPR001715; Calponin-like.
 DR InterPro; IPR001930; Granin.
 DR InterPro; IPR001451; Hexapep_transf.
 DR InterPro; IPR002017; Spectrin.
 DR Pfam; PF00307; CH; 2.
 DR SMART; SM00033; CH; 2.
 DR PROSITE; PS00019; ACTININ 1; 1.
 DR PROSITE; PS00020; ACTININ 2; 1.
 DR PROSITE; PS50021; CH; 2.
 DR PROSITE; PS00422; GRANINS 1; 1.
 DR PROSITE; PS00101; HEXAPEP-TRANSFERASES; 1.
 SQ SEQUENCE 1738 AA; 204427 MW; 577A99D2EC79AF5C CRC64;

Alignment Scores:
 Pred. No.: 1.01e-11 Length: 1738

Score: 318.50 Matches: 180
 Percent Similarity: 37.86% Conservative: 146
 Best Local Similarity: 20.91% Mismatches: 244
 Query Match: 9.36% Indels: 291
 DB: 5 Gaps: 33
 US-09-602-362E-15 (1-2030) x O76329 (1-1738)
 QY 69 TTGAATTGATGACATGCAAACTTTCAAGCAGAGCCTCCGAGAGCCATCTGCCTTC 128
 Db 585 LeuGlnLeuGlnAspGlnGlnAspIleLys-----GluLysGluPheGlnPhe 600
 QY 129 GAGCCT-----GCCATTGAAATGCAA 149
 Db 501 GluLysGlnGlnLeuLeuSerGlnIleAspSerIleThrThrAsnIleGlnGluTyrGln 620
 QY 150 AAGTCTGTTCCAAATAAAGCCTTGGAAATGAAGAATGAACAACATTCAGACAGATGAG 209
 Db 621 AspLysPheAsnAsnLeuGlnGlnPheAsnThrGlnGlnThrLeuAsnGlnGlnGlu 640
 QY 210 ATA-----CTCCCATCAGAAATCCAAACAA-----AAGGACTATCAAGAAAGTTCTTGG 257
 Db 641 ThrHisArgLeuThrGlnGlnLeuTyrGlnIleAsnThrAspTyrAsnGluLysGlnThr 660
 QY 258 GAT-----TCTGAGAGTCTCTGTGACTGT 284
 Db 661 GlnLeuGlnSerGluIleLysAspAsnGlnThrIleAsnGluGlnLeuAsnLysGlnLeu 680
 QY 285 TCACAGAGGATGTGTTTACCCAGGCTACACATCAAAAGAA-----ATAGAT 335
 Db 681 SerGluLysAspLysGluIleGluLysLeuSerAsnGlnGlnGlnGlnGlnAspGlu 700
 QY 336 AAAATAAATGGAATAATAGAAAGTCTCTGATATGATGTTTCTGAAGGTCCTGTC 395
 Db 701 LysIleAsnAsnLeuLeuLeuGluLysGluLysAspCysLeuIleGlu----- 717
 QY 396 AGAATGAAGTTCTATTCACAACTAAAGCCTTAGAA----- 431
 Db 718 -----ArgIleAsnGlnGlnLeuLeuLeuAsnIleAspLeuAsnSerLysTyr 733
 QY 432 -----TTGATGGACATGCAAACTTTCAAGCAGAGCCTCCGAGAGCCATCTGCC 482
 Db 734 GlnGlnLeuLeuLeuGluPheGluAsnPheLys----- 744
 QY 483 TTCAGCCTGCCATTGAATGCAAACTCTGTGTTCCAAATAAAGCCTTGAATGCAAGAT 542
 Db 745 -----LeuAsnSerSerLysGluLysGluAsnGlnLeuAsnGlnLeuGlnSer 760
 QY 543 GAACAAACATTCAGAGCAGATCAGATGTTCCCT----- 575
 Db 761 LysGlnAspGluArgPheAsnGlnLeuAsnAspGluLysGluLysGlnLeu 780
 QY 576 -----TCAGAAATCAAAACAAAGAGGTTGAAGAAATCTTGG 614
 Db 781 GlnSerIleGluAspGluPheAsnGlnTyrLysGlnGlnGlnLeuSerSerAsnSerAsn 800
 QY 615 GATTCAGAGTCTCCGTGAGACTGTTTCACAGAGGATGTGTGTACCCAGGCTACA 674
 Db 801 IleAspGlnGlnLeuGlnSerThrIleIleGlu-----LeuSerGluLeuLys 816
 QY 675 CATCAAAAGAAATGAGT-----AAAATAAGTGGAAATTAGAAGATTCAATAGCCTA 728
 Db 817 GluGlnLysGluLeuAsnAspSerLysLeuIleGluLysGluLysGlnLeuGlnLeu 836
 QY 729 TCAAAATCTTGGATACAGTTTCATCTTGTGAAGACCAAGGAGGAACTTCAAAAGATCAC 788
 Db 837 GlnGlnPheAspGlnLeuAsn-----GluLysAsnGlnLysAspHis 851
 QY 789 TGTGAACAACGTACAGGAAATAATGCAAAATGAAAGAGATTTTGTGTACTGAAAGAG 848
 Db 852 GlnAspGln-----LeuGluLeuLeuGluLysGlnLeuLysGlnLeuGlnGln 867
 QY 849 AAACCTGTGAGAGCAAGAAATAAATCAAGTTAGAACCAAAAGTTAAATGG--- 905

RA Strausberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC050811; AAH50811.1; --
KW Hypothetical protein.
SQ SEQUENCE 520 AA, 61439 MW, 191P1C27B65ABE42 CRC64;

Alignment Scores:

Pred. No.: 2,09e-11 Length: 520
Score: 314.00 Matches: 115
Percent Similarity: 46.04% Conservative: 94
Best Local Similarity: 25.33% Mismatches: 147
Query Match: 9.23% Indels: 98
DB: 11 Gaps: 13

US-09-602-362E-15 (1-2030) x Q80W27 (1-520)

```
QY 603 GAAATCTCTGGAGTCTCGAGAGTCTCCGTGAGACTGTTTCACAG-----AAGGATGTG 656
DB 2 GluAspValGlyAspLeuThrGlnSerSerGluThrAlaSerGluTyGlyArgProPro 21
QY 657 TGTGTACCAAGGCTACACATCAAAAGAAATGGATAAATAGTGGAAATTTAGAGAT 716
DB 22 CysProAsnTyrGluAlaIleLeuArgLeuIleGluArgLeuSerLeuGluCysLysAsp 41
QY 717 TCAACTAGCTATCAAAATCTTGATACAGTTCATCTTGTAAGAGCAAGGCACTT 776
DB 42 SerGlySerLeuLeuLysIleGlnAsnAlaIleHisSerTyrLysArgLeuIleGlnLeu 61
QY 777 CAAAAGATCACTGTGAACAACGTACAGGAAATATGGAACAAATGAAAGAAAGTTTGT 836
DB 62 LysThrGlyHisCysLysValLeuLeuGluLysLeuGlnThrValGluAsnGluValAsn 81
QY 837 GTACTGAAAGAAACTGTCAAGCAAGCAAGAAATAAATACATCACATTAGAGCAACAAA 896
DB 82 GlyLeuArgLysLeuAlaAspAla-----AlaArgGluGln 94
QY 897 GTTAAATGGCAACAGAGCTCGCAGTGTGAGATTGACTTTAAACCAAGCAAGAGAGAG 956
DB 95 LeuGlnGlnGlyGlnGluLeuCysAsnValArgPheThrLeuLysGlnGluGluLys 114
QY 957 AGAAGAAATCGCATATATTAATGAATAAATAGGAAGATTTAGGAAGATTCGAAGAG 1016
DB 115 ArgLysAsnValValTrpValTyrAspLysMetLysGluGlnLeuArgLysGluAsp 134
QY 1017 CAGCATGAGGAAGTGTAGAAAGTGAACCAACTGAAACGGCTCTCAGATACAAAT 1076
DB 135 GlnTyrAsnLysGluValLysMetLysGlnLysLeuGluIleArgValArgGluLeuAsp 154
QY 1077 ATAGATTCGAAGAGTGTAGAAAGTAAATTTGAATCAGTTTCTCACACTCATGAAATGAA 1136
DB 155 MetAspLeuLysThrValArgHisAsnLeuAsnThrVal----- 167
QY 1137 AATTATCTCTTACATGAAATTCATGTTGAAAGAAATTTGCCATGCTAAACCTGAA 1196
DB 168 -----LeuGluGluArgAsnCysIleGluArgGln----- 177
QY 1197 ATAGCCACACTGAAACCAACTACCAAGAAAGAAATAAATTTTCTGAGGACATTAAG 1256
DB 178 -----LeuPheGlnGluGlnAsnThrArgIleIleGlnAsp--Glu 190
QY 1257 ATTTTAAAGAAAGAAATCTGACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 1316
DB 191 IleLeu---AlaAsnHisLeuHisLysGlnLysGluLeuLysMetAlaGlnGluLysIle 209
QY 1317 ACTAAAAGGCGATCTCAATATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGAACAAATG 1376
DB 210 -----SerSerGlnLeuGlnGlnAlaProAspGlnHisThrGln 222
QY 1377 CTCACCTTCTAAATTTGAAGAAACCAAGCAAGAAATCTAGAGCGAGAAATTTGAATCA 1436
DB 223 CysThrGluLysMetArgAspCysIleGlnLysLeuGlyLeuGlu----- 237
QY 1437 CACCATCTTGTGCTTCTGCTGTACAGACCATGATCAAAATTTGTGACATCAAGAAAA 1496
```

```
DB 238 -----AsnHisLysLeuLysGluThrAlaArgLys 247
QY 1497 AGTCAGAACACCTGCTTCCACATTGAGGAGATGCTTTGTTGCAAGAAAAATGAATGTT 1556
DB 248 GlnAlaGluLys-IleGluGlnLeuLysAsnLeu----- 260
QY 1557 GATGTGAGTAGTACCGATATATTAACAAATAGAGTGCTCCATCAACCACTTTCTCAAGCTCA 1616
DB 261 -----SerGlyValLeuSerAspLeuThrGluLysLe 272
QY 1617 AAGGAATCCANAAGCCTAAAAATTAATCTCAATTATGCGAGGAGATGCTCTTAAGAGAAAA 1676
DB 272 uGluSerAlaSerAlaLysTyrLeuHisLeuTyrGluGlnAsnGlnLeuLeuArgGln 292
QY 1677 TACATTGGTTTCAGGAACATGCACAAGAGACCAACGTCGAAACACAGTGTCAATGAAGG 1736
DB 292 uLeuLeuSerMetArgThr-IleGlnArg-----LysCysGlu---LysL 306
QY 1737 AAGCTGAACATGCTNTCAAANGCAACAGATNATGTGAC----- 1777
DB 306 euLysLysHisLysLysLeuGluGlnGluValSerLeuArgSerHisMetLysA 326
QY 1778 -----AAACACTGANCACAGAGTCTC 1802
DB 326 xGAsnMetIleGluCysGlyGlnIleGluGlnTyrLysTyrGluIleGluGluLysThrL 346
QY 1803 TAGATCAGAAATTTTCACTCAACAAAGCAAAATATGTGGCTTCCACAGCAATTAGTTC 1862
DB 346 ysGlnGluValGlyLysLeuLysGlnValAsnLeuPheLeuGlnThrGlnAlaAla 366
QY 1863 ATGCA---CATAANGAAGCTGACAAACAAAGCAAGAT 1897
DB 366 yrgLysAspLysLeuGluLysLeuArgGlnLysGlnAsn 378
RESULT 15
Q15694 PRELIMINARY; PRT; 381 AA.
ID Q15694 AC Q15694 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Protein immuno-reactive with anti-PTH polyclonal antibodies
DE (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBTaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96187433; PubMed=8608414;
RA Kumar R., Haugen J.D., Wieben E.D., Londowski J.M., Cai Q.;
RT "Inhibitors of renal epithelial phosphate transport in tumor-induced
RT osteomalacia and uremia."
RL Proc. Assoc. Am. Physicians 107:296-305(1995).
DR EMBL; U28831; AAB02177.1; -.
DR GO; GO:0016021; C:integral to membrane; NAS.
DR GO; GO:0008200; F:ion channel inhibitor activity; NAS.
DR GO; GO:0030319; P:di-, tri-valent inorganic anion homeostasis; NAS.
FT NON_TER 1
FT NON_TER 381
SQ SEQUENCE 381 AA; 42574 MW; BPE144AB3A21C6DD CRC64;
Alignment Scores:
Pred. No.: 1.08e-10 Length: 381
Score: 302.50 Matches: 99
Percent Similarity: 44.29% Conservative: 60
Best Local Similarity: 27.58% Mismatches: 153
Query Match: 8.89% Indels: 47
DB: 4 Gaps: 7
US-09-602-362E-15 (1-2030) x Q15694 (1-381)
```

Search completed: July 15, 2004, 09:24:01
Job time : 142.318 secs

```
QY 12 AAAGATGCTTCTGGAAGGCTAACTCGGGAATGAAAGTTTCTATTCCAACTAAAGCCTTA 71
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28 LysGlnProAlaLeuLysAlaThrSerAspLysGluAsnSerValProAsnMetAlaThr 47
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 72 GAATTGATGACATGCAAACTTTCAAAGCAGAGCCTCCCGAGAGCCATCGCTTCGAG 131
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48 GluThrLysAspGluGlnIleSerGlyThrValSerSerGlnLysGlnProAlaLeuLys 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 132 CCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCCTTGAATTGAAGATGAACAA 191
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 AlaThrSerAspLysLysAspSerValSerAsnIleProThrGluIleLysAspGlyGln 87
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 192 ACATTGAGACCATGAGATACCTCCCATCAGATCCAAACAAAGAGACTATGAAGNAAGT 251
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88 -----GlnSerGlyThrValSerSerGlnLysGlnProAlaTrpLysAlaThr 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 252 TCTTGGGATTCCTGAGAGTCTCTGTGAG-----TCTCCT 365
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104 SerValLysLysAspSerValSerAsnIleAlaThrGluIleLysAspGlyGlnIleArg 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 279 ---ACTGTTTCACAGAGAGATGTGTTTACCCAGGCTACACATCAAAAAGAAATAGAT 335
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124 GlyThrValSerSerGlnArgGlnProAlaLeuLysAlaThrGlyAsp---GluLysAsp 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 336 AAAATAAATGCAAAATTAGAAGAG-----TCTCCT 365
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 SerValSerAsnIleAlaArgGluIleLysAspGlyGluLysSerGlyThrValSerPro 162
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 366 GATAATGATGTTTCTGAGGCTCCCTCGAGATGAAAGTTTCTATTCCAACTAAAGCC 425
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163 GlnLysGlnSerAlaGlnLysValIlePheLysLysLysValSerLeuLeuAsnIleAla 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 426 TTAGAATTGATGACATGCAAACTTTCAAAGCAGAGCCTCCCGAGAGCCATCTGCCTTC 485
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 ThrArgIleThrGlyGlyTrpLysSerGlyThrGluTrpProGluAsnLeuProThrLeu 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 486 GAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCCTTGAATTGAAGATGAA 545
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 LysAlaThrIleGluAsnLysAsnSerValLeuAsnThrAlaThrLysMetLysAspVal 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 546 CAAACATTGAGCACATCAG-----ATGTTCCCTTCAGAAATCAAAACAAAGAGTT 599
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 223 GlnThrSerThrProGluGlnAspLeuGluMetAlaSerGluGlyGlnLysArgLeu 242
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 600 GAAGAAAATTCCTGGGATTCCTGAGACTCTCCGTGAGACTGTTTTCACAGAG----- 650
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 243 GluGluTy-GluAsnAsnGlnProGlnValLysAsnGlnIleHisSerArgAspAspLeu 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 651 ---GATGTGTGTACCCAGGCTACACATCAAAAGAA----- 686
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 263 AspAspIleIleGlnSerSerGlnThrValSerGluAspGlyAspSerLeuCysAsn 282
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 687 -----ATGGATAAAATAAGTGGAAATATAGAGATTCAACTAGC 725
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 283 CysLysAsnValIleLeuLeuIleAspGlnHisGluMetLysCysLysAspCysValHis 302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 726 CTATCAAAATCTTGATACAGTTCACTTCTGTGAAGAGCAAGGGAATTCAAAAAGAT 785
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 303 LeuLeuLysIleLysThrPheCysLeuCysLysArgLeuThrGluLeuLysAspAsn 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 786 CACTGTGAACAAGTACAGGAAATAAGCAAAATGMAAANGAAGTTTGTGTACTGAAA 845
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 HisCysGluGlnLeuArgValLysIleArgLysLeuLysAsnLysAlaSerValLeuGln 342
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 846 AAGAACTGTGAGAACCAAGAAATAAATACAGTTAGAGAACCAAAAAGTTAAATGG 905
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 343 LysArgSerGluLysGluGluIleLysSerGlnLeuLysHisGluThrLeuGluLeu 362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 906 GACAAAGACTCTGACGTGTGAGATTGACTTTAAACCRAGAGAGAGAGAGAGA 962
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 363 GluLysGluLeuCysSerLeuArgPheAlaIleGlnGlnGluLysLysLys 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 08:45:37 ; Search time 17.1613 Seconds

(without alignments)
12318.669 Million cell updates/sec

Title: US-09-602-362E-15

Perfect score: 3403
Sequence: 1 ctcctgcgttaagatgggt.....gaaaaaaaaaaaaaaaaa 2030

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p,model_p/US09602362/runat_15072004_093625_21988/app_query.fasta_1.10325
-Q/cgn2_1/USPTO.spool_p/US09602362/runat_15072004_093625_21988/app_query.fasta_1.10325
-DB=SwissProt_42 -QFMT=fastan -SUFFIX=resp -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=plo -NORM=ext -HRAPIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09602362 @CGN_1_1_134 @runat_15072004_093625_21988 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|------------|--------------------|
| 1 | 443.5 | 13.0 | 992 | AN18 HUMAN | Q81v16 homo sapien |
| 2 | 280.5 | 8.2 | 1875 | MLP1 YEAST | Q02455 saccharomyc |
| 3 | 280 | 8.2 | 2116 | MY52 DICDI | P08799 dictyosteli |
| 4 | 279 | 8.2 | 2017 | MYGN DROME | Q99323 drosophila |
| 5 | 278.5 | 8.2 | 3911 | AKA9 HUMAN | Q99996 h-a-kinase |
| 6 | 271.5 | 8.0 | 2230 | COM4 HUMAN | Q13439 homo sapien |
| 7 | 269 | 7.9 | 1979 | TRIA HUMAN | Q15643 homo sapien |
| 8 | 266.5 | 7.8 | 2663 | CENE HUMAN | Q02224 homo sapien |
| 9 | 266 | 7.8 | 944 | NUP1 YEAST | P32380 saccharomyc |
| 10 | 262.5 | 7.7 | 1690 | C190 DROME | Q9vj55 drosophila |
| 11 | 262 | 7.7 | 1961 | MYH9 RAT | P62812 rattus norv |
| 12 | 261.5 | 7.7 | 1972 | MYH8 RABIT | P35748 oryctolagus |
| 13 | 260.5 | 7.7 | 1411 | EEAL HUMAN | Q15075 homo sapien |
| 14 | 258 | 7.6 | 1972 | MYH8 MOUSE | C08638 mus musculu |
| 15 | 256.5 | 7.5 | 1790 | USO1 YEAST | P25386 saccharomyc |
| 16 | 255 | 7.5 | 1978 | MYH8 CHICK | P10587 gallus gall |
| 17 | 254 | 7.5 | 2245 | MYGJ DICDI | P54697 dictyosteli |
| 18 | 253 | 7.4 | 1939 | MYH6_MESAU | P13539 mesocricetu |

| | | | | | | |
|----|-------|-----|------|---|------------|--------------------|
| 19 | 252.5 | 7.4 | 1960 | 1 | MYH9 HUMAN | P35579 homo sapien |
| 20 | 251.5 | 7.4 | 2442 | 1 | CEP2 HUMAN | Q9bv73 homo sapien |
| 21 | 250.5 | 7.4 | 1939 | 1 | MYH6 HUMAN | P13533 homo sapien |
| 22 | 250 | 7.3 | 790 | 1 | EEAL MOUSE | Q8b166 mus musculu |
| 23 | 249 | 7.3 | 1938 | 1 | MYH6 RAT | P02563 rattus norv |
| 24 | 247 | 7.3 | 1938 | 1 | MYH6 MOUSE | Q02566 mus musculu |
| 25 | 247 | 7.3 | 1976 | 1 | MYH8 HUMAN | P35580 homo sapien |
| 26 | 247 | 7.3 | 2238 | 1 | COM4 MOUSE | Q91yw5 mus musculu |
| 27 | 247 | 7.3 | 3210 | 1 | CENF HUMAN | P49454 homo sapien |
| 28 | 247 | 7.3 | 3259 | 1 | GOBI HUMAN | Q14789 homo sapien |
| 29 | 246 | 7.2 | 1976 | 1 | MYH8 BOVIN | Q27991 bos taurus |
| 30 | 244 | 7.2 | 879 | 1 | RA50 SULTO | Q96vr5 sulfolobus |
| 31 | 244 | 7.2 | 1130 | 1 | VL17 CAEEL | Q11102 caenorhabdi |
| 32 | 243.5 | 7.2 | 1935 | 1 | MYH7 HUMAN | P12983 homo sapien |
| 33 | 243 | 7.1 | 1364 | 1 | KTNI_CHICK | Q90831 gallus gall |
| 34 | 243 | 7.1 | 1935 | 1 | MYSS CYPCA | C90339 cyprinus ca |
| 35 | 241.5 | 7.1 | 1360 | 1 | CING XENLA | Q9ptd7 xenopus lae |
| 36 | 241 | 7.1 | 997 | 1 | SCPI RAT | Q03410 rattus norv |
| 37 | 240.5 | 7.1 | 1427 | 1 | REST HUMAN | P30622 homo sapien |
| 38 | 240 | 7.1 | 1020 | 1 | CF60 HUMAN | Q8nb25 homo sapien |
| 39 | 240 | 7.1 | 1972 | 1 | MYH8 HUMAN | P35749 homo sapien |
| 40 | 240 | 7.1 | 1976 | 1 | MYH8 RAT | Q9jito rattus norv |
| 41 | 239.5 | 7.0 | 993 | 1 | SCPI MOUSE | Q62209 mus musculu |
| 42 | 238.5 | 7.0 | 1163 | 1 | SBCC CLOAB | Q97fk1 clostridium |
| 43 | 238 | 7.0 | 1938 | 1 | MYHD HUMAN | Q9ukx3 homo sapien |
| 44 | 236.5 | 6.9 | 978 | 1 | RA50 AQUAE | C67124 aquifex aeo |
| 45 | 236.5 | 6.9 | 1935 | 1 | MYH7 RAT | P02564 rattus norv |

ALIGNMENTS

RESULT 1

AN18 HUMAN STANDARD; PRT; 992 AA.
ID Q81v16; Q7Z468;
AC 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ankyrin repeat domain protein 18A.
GN ANKRD18A OR KIAA2015.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Nagase T., Kikuno R., Ohara O.;
RT "The nucleotide sequence of a long cDNA clone isolated from human.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 1-349 FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=22389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marasina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -/- SIMILARITY: Contains 5 ANK repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AB095935; BAC23111.1; ALT INIT.
DR EMBL; BC056266; AAH56266.1; ALT_TERM.
DR Genbank; HGNC:23643; ANKRD18A.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 6.
DR PRINTS; PR01415; ANKYRN.
DR SMART; SM00248; ANK; 6.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
KW ANK repeat; Repeat.
FT REPEAT 67 96 ANK 1.
FT REPEAT 100 129 ANK 2.
FT REPEAT 133 162 ANK 3.
FT REPEAT 166 195 ANK 4.
FT REPEAT 199 228 ANK 5.
FT CONFLICT 130 130 K -> E (IN REF. 2).
FT CONFLICT 288 288 MISSING (IN REF. 2).
SQ SEQUENCE 992 AA; 115664 MW; B9B288F087340D9A CRC64;

Alignment Scores:
Pred. No.: 1,37e-16 Length: 992
Score: 443.50 Matches: 131
Percent Similarity: 50.84% Conservative: 80
Best Local Similarity: 31.57% Mismatches: 164
Query Match: 13.03% Indels: 40
DB: 1 Gaps: 6

US-09-602-362E-15 (1-2030) x AN18_HUMAN (1-992)
QY 815 ACAATGAAAGAAAGCTTTGTGTACTGTAAGAAAGAACTGTGAGA----- 859
Db ThrAlaGluAspTyrAlaLeuCysSerAspLeuArgSerIleArgGlnGlnIleLeuGlu 254
QY 860 -----AGCAAGAAATAAATACAGTTAGACACCAAAAGTTAAATGGACACAGA 913
Db HisLysAsnLysMet-LeuLysAsnHisLeuArgAsnAsp----- 267
QY 914 GCTCTGCAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAAATGCCGATAT 973
Db -----AsnGlnGluThrAlaAlaMetLysProAla----- 277
QY 974 ATTAATGAAATAATTAGGAGAGAAATAGGAGAGAAATCGAGAGACGATAGGAAAGATT 1033
Db -----AsnLeuLysLysArgLysGluArgAlaLysAlaGluHisAsnLeuLysValAlaLe 296
QY 1034 AGAAGTGAACCAACACTTGAACAGCTCTCAGAAATACAGATATAGAAATGAGAGTGT 1093
Db -----ArgLeuGlnArgSe 306
QY 1094 AGAAATTAATTGAATCAGCTTTCTCAGCTCATGAAATGAAATATCTCTTACATGA 1153
Db -----AsnLeuLysGlnProGlnAspSerGlnSerTyrGlyLysLysLysAspAlaMetTyrG 326
QY 1154 AAATTCATGTTGAAAAGCAATGCCATGCTAAAGCTGAAATAGCCACACTGAAACA 1213
Db -----AsnLeuLysLysAspIleAlaMetLeuLysGluGluLeuTyrAlaIleLysAs 346
QY 1214 CCAATACCAAGGAAAGAAATAATACCTTTGAGACATTAAGATTTTAAAGAAAGAA 1273
Db -----AspSerLeuArgLysGluLysTyrIleGlnGluIleLysSerIleThrGluIleAs 366
QY 1274 TGCTGAATTCAGATGACCTTAAACTGAACAGAGGAATCATTAATGAAAGGCCATCTCA 1333

Db nAlaAsnPhleGluLysSerValArgLeuAsnGluLysMetIleThrLysThrValAlaAr 386
QY 1334 ATATAGTGGCAGCTTAAAGTTCTGATAGCTAGACACACATCTCACTCTCTAAATG-- 1391
Db gTyrSerGlnGlnLeuAsnAspLeuLysAlaGluAsnAlaArgLeuAsnSerGluLeuG 406
QY 1392 -AAGAAAAACAAGACACAAAGAAATATCTAGAGGAGAGAAATGAATCAACCATCTAGAT 1450
Db uLysGluLysHisAsnLysGluArgLeuGluAlaGluValGluSerLeuHisSerLe 426
QY 1451 GCCTTCGTGTACAGACCATCATCAATTGTGACATCAAGAAAAAGTCAAGAACTGC 1510
Db uAlaThrAlaIleAsnGluTyrAsnGluIleVal---GluArgLysAspLeuGluVal 445
QY 1511 TTTCCACATTCGAGGAGATGCTTTGTCACAAAGAAATGAATGTTGATGTGATGATGATAC 1570
Db LLeuTyrArgAlaAspAspValSerArgHisGluLysMetGlySerAsnIleSerGlnLe 465
QY 1571 CGATATATAACAATGAGGTGCTCCATCAACCACTTTCTGAAGCTCAAGAAATCCANAA 1630
Db u-ThrAspLysAsnGluLeuThrGluGlnValHisLysAlaArgValLysPheAsnT 485
QY 1631 GCCTAAAAATTAATCTCAATTATGAGGAGATGCTCTAAGAGAAATATACATGTTTCAG 1690
Db hLeuLysGlyLysLeuArgGluThrArgAspAlaLeuArgGluLysThrLeuAlaLeuG 505
QY 1691 GAACATGCACAAAGACACCAACCTGAAACACAGTGTCAATGAAGAAAGCTGAACATG 1750
Db lYSer-ValGlnLeuAspLeuArgGlnAlaGlnHisArgIleLysGluMetLysGlnMet 524
QY 1751 TATCAANCGAACCAAGATNATGTGAACAAACACACTGANCAGCAGGAGTCTCTAGATCAG 1810
Db HisProAsnGlyGluAlaLysGluSerGlnSerIleGlyLysGlnAsnSerLeuGluGlu 544
QY 1811 AAATTTTTCACACTCAAAAGCAAAATATGTGCTTCAACAGCAATTAGTTCATGCACAT 1870
Db ArgIleArgGlnGlnGluLeuGluAsnLeuLeuGluArgGlnLeuGluAspAlaArg 564
QY 1871 AANGAAGCTGCACACAAAGCAAGATACAAATGATGATNTTCATNTCTTGAGAGGAAAT 1930
Db LysGluGlyAspAsnLysGluIleValIleAsnIleHisArgAspCysLeuGluAsnGly 584
QY 1931 GCNCATCATCTCTAAAGAGAGAAATAGAGAGATATTTNATACNATAACCATTTAAAA 1990
Db LysGluAspLeuLeuGluGluArgAsnLysGluLeuMetLysGluTyrAsnTyrLeuLys 604
QY 1991 AACCGGTATATTTCAATATGGAAGAAAAAANAAAAA 2029
Db GlulysLeuLeuGlnCysGluLysGluLysAlaGluArg 617

RESULT 2
MLP1_YEAST STANDARD; PRT; 1875 AA.
ID MLP1_YEAST
AC Q02455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-like protein MLP1.
GN MLP1 OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=33247549; PubMed=8483450;
RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Gen. Genet. 237:359-369 (1993).
RN [2]
RP SEQUENCE FROM N.A.

```

RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey P., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
RL new open reading frames.";
RL Yeast 9:1349-1354(1993).
CC -!- FUNCTION: Myosin-like protein that is probably involved in DNA
CC repair.
CC -!- SIMILARITY: SOME TO THE TPR ONCOGENE.
CC -!- CAUTION: Ref.2 misquotes the gene name as "MPL1".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L01992; AAA34783.1; -.
DR EMBL; X73541; CAA51948.1; -.
DR EMBL; Z28320; CAA82174.1; -.
DR PIR; S38173; S38173.
DR GerMoliner; 140074; -.
DR SGD; S0001803; MPL1.
DR GO; GO:0005635; C:nuclear membrane; IDA.
DR GO; GO:0005654; C:nucleoplasm; IDA.
DR GO; GO:0006006; P:protein-nucleus import; IDA.
KW Coiled coil; DNA repair.
FT DOMAIN 59 487 COILED COIL (POTENTIAL).
FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
FT CONFLICT 301 301 R -> A (IN REF. 1).
SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C9066867 CRC64;

Alignment Scores:
Pred. No.: 4,36e-08 Length: 1875
Score: 280.50 Matches: 181
Percent Similarity: 40.05% Conservative: 145
Best Local Similarity: 22.24% Mismatches: 277
Query Match: 8.24% Indels: 211
Dbs: 1 Gaps: 39

US-09-602-362E-15 (1-2030) x MPL1_YEAST (1-1875)
QY 33 AACTCGGAATGAAGTTCTATT---CCACTAAAGCCTTAGAATTGATGCACATGCAA 89
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
737 SerCysLysSerLysLeuSerIleValGluThrGluLeuLeuAsnLeuLysGluGluGln 756
QY 90 ACTTTCAAGCAGAG-----CCTCCCGAG 113
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
757 LysLeuArgValHisLeuGluLysAsnLeuLysGlnGluLeuAsnLysLeuSerProGlu 776
QY 114 AAGCCATCTCGCTTCGAGCTCCCATTCGAAATGCAAAAGTCTGTTCCAAATAAGCCTTG 173
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
777 LysAspSerLeuArgIleMetValThrGlnLeuGln-----ThrLeu 790
QY 174 GAATTGAAGAAATGAACAACATG-----AGACGACAT 206
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
791 GlnLysGluArgGluAspLeuLeuGluThrArgLysSerCysGlnLysLysIleAsp 810
QY 207 GAGATA-----CTCCCATCAGAAATCAACAAAGGACTAT----- 242
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
811 GluLeuGluAspAlaLeuSerGluLeuLysLysGluThrSerGlnLysAspHisIle 830
QY 243 -----GAGAAAGTTCT-----TGG----- 257
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
831 LysGlnLeuGluAspAsnAsnSerAsnIleGluTrpTyrGlnAsnLysIleGluAla 850
QY 258 -----GATTCTCAGAGTCTCTGTGAGACTGTTTCACAGAGAGGATGTGTGTTTACC 308
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
851 LeuLysLysAspTyrGluSerValIleThrSerValAspSerLysGlnThrAspIleGlu 870

```

```

QY 309 AAGGCTACACAT-----CAAAAGAAATA-----CATAAATAAAT----- 344
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
871 LysLeuGluTyrLysValLysSerLeuGluLysGluIleGluAspLysIleArgLeu 890
QY 345 -----GGAAATTAAGAGCTCTCTGATAATGATGGTTTCTGAAGCTCCCTGC 395
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
891 HisThrTyrAsnValMetAspGluThrIleAsnAspSerLeuArgLysGluLeuGlu 910
QY 396 AGAATGAAGATTCTATTCCA-----ACTAAAGCCTTAGAATTGATGCACATG--- 443
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
911 LysSerLysIleAsnLeuThrAspAlaTyrSerGlnIleLysGluTyrLysAspLeuTyr 930
QY 444 -----CAACTTTCAAAGCAGAGCTCCCGAAGCAGCATCTCCCTTCGAGCCT 491
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
931 GluThrThrSerGlnSerLeuGlnThrAsnSerLysLeuAspGluSerPheLysAsp 950
QY 492 GCCATTGAATGCAAAAGCTGTTCCAAATAAGCCTTCGAAATTGAAGAAATGAACAAACA 551
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
951 PheThrAsnGlnIleLysAsnLeuThrAspGluLysThrSerLeuGluAspLysIleSer 970
QY 552 TTGAGAGCAGATCAGATGTTCCCTTCAGAAATCAAAA-----CAAAGAGGTTGAA 602
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
971 LeuLysGluGlnMetPheAsnLeuAsnAsnGluLeuAspLeuGlnLysLysGlyMet 990
QY 603 GAAATTTCTGGGATTCTGAGACTCTCCGTGAGACTGTTTCACAG-----AAGGAT 653
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
991 GluLysGluLysAlaAspPheLysLysArgIleSerIleLeuGlnAsnAsnLysGlu 1010
QY 654 GTGTGTGTACCCAGGCTCACATCAAAAGAAATGGAATAAATAGTGGAAAATTAGAA 713
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1011 ValGluAlaValLysSerGluTyrGluSerLysLeuSerLysIleGlnAsnAspLeuAsp 1030
QY 714 GATTCAACT----- 722
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1031 GlnGlnThrIleTyrAlaAsnThrAlaGlnAsnAsnTyrGluGlnGluLeuGlnLysHis 1050
QY 723 -----AGCCTATCAAAAATCTTCGATACACATCTTCATCTCTGTGAA----- 761
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1051 AlaAspValSerLysThrIleSerGluLeuArgGluGlnLeuHisThrTyrLysGlyGln 1070
QY 762 ---AGAGCAAGGGAACCTTCAAAAGATCACTGTGAACAAGTACAGGAAAAATGGAACAA 818
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1071 ValLysThrLeuAsnLeuSerArgAspGlnLeuGlu-----AsnAlaLeuLysGlu 1087
QY 819 ATGAAAAGAGTTTGTGTACTGAAAAGAAAATGTGCAAGAACAAAGAAATA---AAA 875
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1088 AsnGluLysSerTrpSerSerGlnLysGluSerLeuLeuGluGlnLeuAspLeuSerAsn 1107
QY 876 TCACAGTTAGAGAAC-----CAAAAGTTAAATGGAACAA----- 911
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1108 SerArgIleGluAspLeuSerSerGlnAsnLysLeuLysLeuTyrAspGlnIleTyr 1127
QY 912 -----GAGCTCTCAGTGTGAGA 929
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1128 ThrAlaAlaAspLysGluValAsnAsnSerThrAsnGlyProGlyLeuAsnIleLeu 1147
QY 930 TTGACTTTAAACCAAGAAAGAGAG-----AAGAGAGAGAAAT 965
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1148 IleThrLeuArgArgGluArgAspIleLeuAspThrLysValThrValAlaGluAsp 1167
QY 966 GCCGATATATAATGAAAAATT----- 989
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1168 AlaLysMetLeuArgGlnLysIleSerLeuMetAspValGluLeuGlnAspAlaArgThr 1187
QY 990 -----AGGGAAGAATTAGGAAGATTCGAAGACGACATAGG 1025
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1188 LysLeuAspAsnSerArgValGluLysGluAsnHisSerSerIleIleGlnGlnHisAsp 1207
QY 1026 AAAGAGTTAAGTGAACAAACAACTTGAACAGCTCTCAGATAACAAGATATAGATTG 1085
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1208 AspileMetGluLysLeuAsnGlnLeu---AsnLeuLeuArgGluSerAsnIleThrLeu 1226

```

```

QY 1086 AGAGTGTGAGAAAGTAATTTGAATCAGGTTTCTCAGCTCATGAAATGAAATATATCTC 1145
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1227 Arg - - - - - AenGluLeuGluAsnAsnAsnLysLysLysGluLeu 1240
QY 1146 TTACATGAAATTCATGTTGAAAGAGAAATTCCTCATGCTAAATCGAATGCCACA 1205
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1241 GlnSerGluLeuAspLysLeuGlnAenValAlaProIleGluSerGluLeuThrAla 1260
QY 1206 CTGAACACCAATACACAGGAAAGGAAATAATCTTTGAGACATTAAGATTTTAAAA 1265
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1261 LeuLysTy-SerMetGlnGluLysGlu - - - - - GlnGluLeuLysLeuAlalys 1276
QY 1266 GAAAGAAATGCTCAATTCAGATGACCCCTAAACTGAAGAGAGATCATTAAGTAAAGG 1325
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1277 Glu - - - - - GluValHisArgTrpLysLysArgSerGlnAspIleLeuGluLysHis 1293
QY 1326 GCATCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGAACACACATGCTCACTTCT 1385
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1294 GluGlnLeuSerSerSerAspTyrgLysLysLeuGluSerGlu - - - - - IleGlu 1309
QY 1386 AAATTGAGGAAAA - - - - - CAAGACAAAGAAATCTAGAGCGAGAAATTCATCACCAC 1442
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1310 AsnLeuLysGluGluLeuGlnAsnLysGluArgGlnGlyAlaGluAlaGluLysPhe 1329
QY 1443 CCTAGACTGGCTTCTGCTGTACAGACCATGATCAATTTGTACATCAAGAAAAAGTCAA 1502
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1330 AsnArgLeuArgGlnAlaGln - - - - - GluArgLeuLysThrSerLysLeuSerGln 1347
QY 1503 GAACCTGCTTTCACATTCAGGAGATGCTTTTTCGAAAGAAATGAATGTTGATGTG 1562
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1348 Asp - - - - - SerLeuThrGluGlnValAsn - - - - - 1355
QY 1563 AGTAGTACCGATATATAAATGAGTGTCTCCATCAACCTTCTGAGCTCAAGGAA 1622
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1356 SerLeuArgAspAla - LysAsn - - - - - ValLeuGluAsnSerLeuSerGluAlaAsnAlaAr 1374
QY 1623 ATCCNAAGCCCTAA - - - - - ATTAATCTCAATATGCGAGAGATGCTCAAGAGAAATAC 1679
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1374 gileGluGluLeuGlnAsnAlaLysValAlaGlnGlyAsnAsnGlnLeu - GluAlaIleA 1394
QY 1680 ATGTTTTCAGGACATGACACAAAGAGACCAACGTCGAAACACAGTGTCAATGAAGGAG 1739
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1394 rgLysLeuGlnGluAspAlaGluLysAlaSerArgGluLeuGlnAlaLysLeuGluGln 1414
QY 1740 CTGACACATGTTTCAAAACGAACAGATATGTGACAAACAC - - - - - ACTGACGACGAG 1796
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1414 erThrThrSerTyrgLysThrIleAsnGlyLeuAsnGluGluIleThrThrLeuLysG 1434
QY 1797 AGTCTCTAGATCAGAAATATTTCACACTACAAGCAAAATATGTGGCTTCACAGCAAT 1856
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1434 luGluIleGlu - - - - - LysGlnArgGlnIleGlnGlnGln 1446
QY 1857 TAGTTCATCATGAAGAAGCTGACACAAAGCAAGATAACAAATTCATTTCA - - - - - 1912
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1446 eu - - - - - GlnAlaThrSerAlaAsnGluGlnAsnAspLeuSerAsnIleValGluSerMetL 1465
QY 1913 - - - - - TTNTCTTCAGAGGAAATGCNCTCATCTTCTTAAAGAGAAA - - - - - AATG 1958
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1465 ysLysSerPheGluGluAspLysIleLysPheIleLysGluLysThrGlnGluValAsnG 1485
QY 1959 AGGAGATATTTTATACNATACCATTTTAAAAAACCAG 1996
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1485 lylsileLeuGluAlaGlnGluArgLeuAsnGlnPro 1497

```

RESULT 3

MYS2 D1CD1

ID MYS2 D1CD1

AC P08759;

DT 01-NOV-1988 (Rel. 09, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Myosin II heavy chain, non muscle.

GN MHCA.

```

OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97092266; PubMed=3540939;
RA Warrick H.M., de Lozame A., Leinward L.A., Spudich J.A.;
RT "Conserved protein domains in a myosin heavy chain gene from Dictyostelium discoideum."
RL Dictyostelium discoideum."
RN [2]
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
RX MEDLINE=90353583; PubMed=2387408;
RA Luck-Vielmeier D., Schleicher M., Grabatin B., Wippler J., Gerisch G.;
RT "Replacement of threonine residues by serine and alanine in a phosphorylatable heavy chain fragment of Dictyostelium myosin II."
RL FEBS Lett. 269:239-243(1990).
RN [3]
RP PHOSPHORYLATION SITES.
RX MEDLINE=88112226; PubMed=2828113;
RA Wagle G., Noegel A., Scheel J., Gerisch G.;
RT "Phosphorylation of threonine residues on cloned fragments of the Dictyostelium myosin heavy chain."
RL FEBS Lett. 227:71-75(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX MEDLINE=95345066; PubMed=7619795;
RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M., Rayment I.;
RT "X-ray structures of the myosin motor domain of Dictyostelium discoideum complexed with MgADP.Berx and MgADP.ALFA-";
RL Biochemistry 34:8960-8972(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX MEDLINE=95345067; PubMed=7619796;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the truncated head of Dictyostelium discoideum myosin to 2.7-A resolution."
RL Biochemistry 34:8973-8981(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=96206189; PubMed=8611530;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II) ADP.vanadate complex of the Dictyostelium discoideum myosin motor domain to 1.9-A resolution."
RL Biochemistry 35:5404-5417(1996).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX MEDLINE=97452580; PubMed=9305951;
RA Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT "X-ray structures of the MgADP, MgATPgammA, and MgAMPNP complexes of the Dictyostelium discoideum myosin motor domain."
RL Biochemistry 36:11619-11628(1997).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=98070605; PubMed=9405148;
RA Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
RT "X-ray crystal structure and solution fluorescence characterization of Mg.2'(3')-O-(N-methylanthraniloyl) nucleotides bound to the Dictyostelium discoideum myosin motor domain."
RL J. Mol. Biol. 274:394-407(1997).
CC -I- FUNCTION: Myosin is a protein that binds to actin and has ATPase activity that is activated by actin.
CC -I- SUBUNIT: Myosin II heavy chain is two-headed. It self-assembles into filaments. Hexamer of 2 heavy chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2 regulatory light chain subunits (MLC-2).
CC -I- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL CORTEX.
CC -I- DOMAIN: Each myosin heavy chain can be split into 1 light

```


QY 249 AGTCTTGGGAT-----TCTGAGAGTCTCTGTGAGACTGTCTTTCACAGAAGGAT 296
 Db 955 GluLeuGluMetLysArgValAsnAspGlyGlnSerAspThrIleSerArgLeuGlu 974
 QY 297 GTGTGTTTACCCAAAGCTACACATCAAAAGAAATAGATAAAATAAATGGAATA--- 353
 Db 975 LysIle-----LysAspGlnLeuGlnLysGluValGluLeuThrGlnSerPheSer 992
 QY 354 GAAGAGTCTCTGATTAATGATGTTTCTGAGAGCTCCCTCGCAGATGAAGTCTTCTATT 413
 Db 993 GluGluSerLysAspLysGlyValLeuGluLysThrArgValArgLeuGlnSerGluLeu 1012
 QY 414 CCAACTAAAGCCTTAGAATGTAGGACATGCAAACTTTCAAAGCAGAG----- 461
 Db 1013 AspAspLeuThrValArgLeuAspSerGluThrLysAspLysSerGluLeuLeuArgGln 1032
 QY 462 -----CCTCCGAGAGCCATCTCCCTTCGAGCTGCC 494
 Db 1033 LysLysLysLeuGluGluLeuLysGlnValGlnGlnAlaLeuAlaGluThrAla 1052
 QY 495 ATTGAATGCAAAAGCTCTGTTCCAATAAAGCCTTG----- 530
 Db 1053 AlaLysLeuAlaGlnGluAlaAlaAsnLysLysLeuGlnGlyGluTyrThrGluLeuAsn 1072
 QY 530 ----- 530
 Db 1073 GluLysPheAsnSerGluValThrAlaArgSerAsnValGluLysSerLysLysThrLeu 1092
 QY 531 -----CAATTGAAGATGAACAACA 551
 Db 1093 GluSerGlnLeuValAlaValAsnAsnGluLeuAspGluGluLysLysAsnArgAspAla 1112
 QY 552 TTGAGA-----GCAGATCAGATGTC 572
 Db 1113 LeuGluLysLysLysLysAlaLeuAspAlaMetLeuGluGluMetLysAspGlnLeuGlu 1132
 QY 573 CCTTCAGATCAAAACAAAG-----AAGTTGAAGAAATCTTTGGGAT 617
 Db 1133 SerThrGlyGlyLysLysSerLeuTyrAspLeuLysValLysGlnGluSer---Asp 1151
 QY 618 TCTGAGAGTCTCGTGAGAGTCTTTCACAGAGGATGTGTGTACCC----- 665
 Db 1152 MetGluAlaLeuArgAsnGlnIleSerGluLeuGlnSerThrIleAlaLysLeuGluLys 1171
 QY 666 ---AAGGCTACATCAAAAGAAATGGATAAAATAGTGGAAATTAAGAGATTCAACT 722
 Db 1172 IleLysSerThrLeuGluGlyGluValAlaArgLeuGlnGlyGluLeuGlu----- 1188
 QY 723 AGCCTATCAAAATCTTGATACAGTTCATTCTTGAAAGAGCAAGGAACTTCAAAA 782
 Db 1189 -----AlaGluGlnLeuAlaLys 1194
 QY 783 GATCACTGTGACACAGCTACAGAAATCGAA-----CAAATGAAAGAAAGTTTGTGTA 839
 Db 1195 SerAsnValGluLysGlnLysLysLysValGluLeuAspLeuGluAspLys----- 1211
 QY 840 CTGAAAAGAACTGTGACAGCAAGCAAGAAATAAATACACATTAGACACCAAAAGTT 899
 Db 1212 ---SerAlaGlnLeuAlaGluGluThrAlaAlaLysGlnAlaLeuAspLysLeuLysLys 1230
 QY 900 AAATGGGAACAAGAGCTCTGCAGTGTGAGATTCACCTTTAAACCAAGCAAGAGAGAGA 959
 Db 1231 LysLeuGluGlnGluLeuSerGluValGlnThrGlnLeuSerGluAlaAsnAsnLysAsn 1250
 QY 960 AGAAATGCCGATATATAATGAAATAATAGGAAAGAAATAGGAAAGATTCGAGAGCAG 1019
 Db 1251 ValAsnSerAspSerThrAsnLysHisLeuGluThrSerPheAsnAsnLeu----- 1267
 QY 1020 CATPAGGAAGAGTTAGAACTGAAACAACACTTGAACAGGCTCTCAGATAACAATATA 1079
 Db 1268 ---LysLeuGluLeuGluAlaGluGlnLysAlaLysGlnAlaLeuGluLysLysArgLeu 1286

QY 1080 GAATTGAAGAGTGTAGAAAGTAATTTGATCATAGGTTTCTCACACTCATGAAATGAAAT 1139
 Db 1287 GlyLeu-----GluSerGluLeuLysHisValAsnGluGlnLeuGluGluLys 1303
 QY 1140 TATCTCTTACATGAAATTTGCATGTTGAAAGAAATGCCCATTGCTAAACCTGGAATA 1199
 Db 1304 LysGlnLysGluSerAsn-----GluLysArgLysValAspLeuGluLysGluVal 1320
 QY 1200 GCCACACTGAAACACCAATACACAGGA----- 1226
 Db 1321 SerGluLeuLysAspGlnIleGluGluGluValAlaSerLysLysAlaValThrGluAla 1340
 QY 1227 -----AAGGAAAT-----AAATACTTTGAGGACATTAAGATT 1259
 Db 1341 LysAsnLysLysGluSerGluLeuAspGluIleLysArgGlnTyrAlaAspValValSer 1360
 QY 1260 TTAAGAAAGAAATGCTGAATCTGAGATGACCTTAAACTGAAGAGGAATCATTAAT 1319
 Db 1361 SerArgAspLysSerValGluGlnLeuLysThrLeuGlnAlaLysAsnGluGluLeuArg 1380
 QY 1320 AAAAGGCACTCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGAACACATGCTC 1379
 Db 1381 AsnThrAlaGluGluAlaGluGlyGlnLeu----- 1390
 QY 1380 ACTTCTTAATTAAGGAAACAAAGCAAGCAAGAAATACTAGAGGACAGAAATGATCACAC 1439
 Db 1391 -----AspArgAlaGluArgSerLysLysLysAlaGluPheAsp----- 1403
 QY 1440 CATCTAGACTGCTTCTGCTGTACCAAGACCATCATCAAAATGTGACATCAAGAAAAAGT 1499
 Db 1404 -----LeuGluGluAlaValLysAsnLeuGluGluThrAlaLysLysValLys 1420
 QY 1500 CAAGAACTCTCTTCCACATGTCAGGAGATGTTGTTGCAAGAAAAATGAATGTTGAT 1559
 Db 1421 AlaGluLysAla-----MetLysLysAlaGluThrAsp 1431
 QY 1560 GTGAGTAGTACCGATATATAACAATGAGGTGCTCCATCAACCACTTTCTGAAGCTCAAAG 1619
 Db 1432 TyrArgSerThr-----LysSer 1437
 QY 1620 GAAATCCANAAGCCTAAAAATTAATCTCAATTATGAGGAGATGCTTAAGAAAAATAC 1679
 Db 1438 GluLeuAspAspAlaLysAsn-----ValSerSerGluGlnTyr 1450
 QY 1680 ATGGTTTTCAGGAACATGCAAAAGAGACCAACGCTGAAACACAGTCTCAATGAAGGAAG 1739
 Db 1451 ValGlnIleLysArgLeuAsnGluGluLeuSerGluLeuArgSerValLeuGluGluAla 1470
 QY 1740 CTGAACACATGTTTCAAAACGAAAGATNATGTGAACAAACACACTGANCACAGGAGT 1799
 Db 1471 AspGluArgCysAsnSerAlaIleLysAlaLysLysThrAlaGluSerAlaLeuGluSer 1490
 QY 1800 CTCTAGATCAGAAATTAATTTCAACTCAAGCAAGCAAAATATGTGCTTCAACAGCAATTAG 1859
 Db 1491 Leu-----LysAspGluIleAspAlaAlaAsnAsn----- 1500
 QY 1860 TTCATGCACATAANGAAAGCTGACAAACAAAGCAAG 1895
 Db 1501 -----AlaLysAlaLysAlaGluArgLysSerLys 1510
 RESULT 4
 MYSN DROME
 ID MYSN DROME STANDARD; PRT; 2017 AA.
 AC Q99323;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Myosin heavy chain, non-muscle (zipper protein) (Myosin II).
 GN ZIP.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.

US-09-602-362E-15 (1-2030) x AKA9_HUMAN (1-3911)

135 GCCATTGAATGCAAAAGCTGTTCCAAATAAAGCCTTGAATTAAGAAATGAA----- 188
112 SerValGluLeuGluSerGluIleSerThrThrAlaAspAspCysSerSerGluValAsn 131
189 -----CAAAACATTGAGACGACGATGAGATA 212
132 GlyCysSerPheValMetArgThrGlyLysProThrAsnLeuLeuArgGluGluPhe 151
213 CTCCCATCAGATCCAAACAAAGGACTATGAGAAAGTCTTGGGATTCTCAGAGTCTC 272
152 GlyValAspSerThrSer-----GluGlnGlyAlaGlnAspSerProThrHis 168
273 TGTGAGACTGTTTCACAGAAAGGATGTGTGTTTACCAAGGCTACACATCAAAAGAAATA 332
169 LeuGluMetMetGluSerGlu-----LeuAlaGlyLysGlnHisGluIle 183
333 GATAAAATTAATGGAATTAAGAGAGTCTCCTGATATGATGGTTTCTGAGGCTCCC 392
184 GluGluLeuAsnArgGluLeuGluGlu----- 192
393 TGCAGAAATGAAAGTCTTCTCACTCACTAAAGCCTTAGAATTGATGACATGCAAACTTTC 452
193 -----MetArgValThrThrGlyThrGluGlyLeuGln-----GlnLeuGlnGluPhe 208
453 AAAGCAGAGCCTCCGAGAGAGCCATCTGCCTCGAGCCTGCCATTCGAATGCAAAAGTCT 512
209 GluAlaAlaIleLysGlnArg-----AspGlyIleIleThrGlnLeuThr 223
513 GTTCCAAATAAAGCCTTGGNAITGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATG 572
224 AlaAsnLeuGlnAlaArgGluLysAspGluThrMetArgGluPheLeuGluLeu 243
573 CTTTCAGAAATCAAAACAAAGAGGTTGAA-----GAAAATCTTGGGATTCTCAGAGTCTC 629
244 ThrGluGlnSerGlnLysLeuGlnIleGlnPheGlnGlnLeuGlnAlaSerGluThrLeu 263
630 CTGAGACTGTT-----TCACAGAAGAGTGTGTGTACCCCAAG-----GCT 671
264 ArgAsnSerThrHisSerThrAlaAlaAspLeuLeuGlnAlaLysGlnGlnIleLeu 283
672 ACACATCAAAAGAAATGATAAATAAGTGAATAAATAGAAATGAGAAATGAGAAATGAGAAAT 716
284 ThrHisGlnGlnGlnLeuGluGlnAspHisLeuLeuGluAspTyrGlnLysLysLys 303
717 -----TCAACTAGCCTATCAAAATCTTGGAT-----ACAGTTCAATCTTGTGAA 761
304 GluAspPheThrMetGlnIleSerPheLeuGlnGlnLysLysLysLysValTyrGluMetGlu 323
762 AGAGAAGGGAACCTTCAAAAGATCAT-----TGT 791
324 GlnAspLysLysValGluAsnSerAsnLysGluGluIleGlnGluLysGluThrIleIle 343
792 GAACAACGCTACAGGAAATGGAACAAATGAAAGAAAGCTTTTGTGTCTGAAAGAA 851
344 GluGluLeuAsnThrLysIleIleGluGluLysLysLysLysLysLysLysLysLys 363
852 CTGTCAAGAGCAAAAGAAATAAATACACAGTTAGAGAACCAAAAGTTAAATGGGAACAA 911
364 LeuThrThrAlaAspLysLeuGlyGluLeuGlnGlnIleValGlnLysAsnGln 383
912 GAGCTCTGAGTGTGAGATTGATTTTAAACCAAGAGAGAGAGAGAAATGCCGAT 971
384 GluIleLysAsnMetLysLeuGluLeuThrAsnSerLysGlnLysGluArgGlnSer--- 402
972 ATATTAAATCAAAATATTAGGAGAGATTAAGAGAAATCGAGAG----- 1016
403 -----SerGluIleLysGlnLeuMetGlyThrValGluGluLeuGlnLysArgAsn 420
1016 ----- 1016

421 HisLysAspSerGlnPheGluThrAspIleValGlnArgMetGluGlnGluThrGlnArg 440
1016 ----- 1016
441 LysLeuGluGlnLeuArgAlaGluLeuAspGluMetTyrGlyGlnGlnIleValGlnMet 460
1017 -----CAGCATAGAAAGAGATTAGAGTG-----AAA 1043
461 LysGlnGluLeuIleArgGlnHisMetAlaGlnMetGluGluMetLysThrArgHisLys 480
1044 CAACAACCTTGAACACAGGCTCTCAGA----- 1067
481 GlyGluMetGluAsnAlaLeuArgSerTyrSerAsnIleThrValAsnGluAspGlnIle 500
1068 -----ATACAAGATATAGAAATTTGAAGAGTGTAGAAATTAATTTGAAT 1109
501 LysLeuMetAsnValAlaIleAsnGluLeuAsnIleLysLeuGlnAspThrAsn---Ser 519
1110 CAGGTTTCTCACACTCATGAAATGAAATTTACTCTTACATGAAATTCATGTTGAAA 1169
520 GlnLysGlnLysLysLysGluGluLeuGlyLeuIleLeuGluLysCysAlaLeuGln 539
1170 AAGAAATTCCTGCTTAAACTGGAATAGCCACACTGAAACACCAATACCCAG----- 1223
540 ArgGlnLeuGluAspLeuValGluLeuSerPheSerArgGluGlnIleGlnArgAla 559
1224 -----GAAAAGGAAATAATATAC----- 1241
560 ArgGlnThrIleAlaGluGlnGluSerLysLeuAsnGluAlaHisLysSerLeuSerThr 579
1242 TTTGAGGACATTAAGATTTTAAAGAAAGAAATGCTGAACTTCAGATGACCTTAAACTG 1301
580 ValGluAspLeuLysAlaGluIleValSerAlaSerGluSerArgLysGluLeuGluLeu 599
1302 AAAGAGAAATCAATTAATAAGGGCATCTCAATATAGTGGCAGCTTAAAGTTCTGATA 1361
600 LysHisGluAla-----GluValThrAsnTyrLysIleLysLeuGluMetLeuGlu 616
1362 GCTGAGAACACAATGCTCACTTCTAAATTTGAAGGAAACAAAGACAAAGAAATA----- 1415
617 LysGluLysAsnAlaValLeuAspArgMetAlaGluSerGlnGluAlaGluGluArg 636
1416 -----CTAGAGGAGAAAT 1430
637 LeuArgThrGlnLeuLeuPheSerHisGluGluGluLeuSerLysLysLysGluAspLeu 656
1431 GAATCACACACATCTAGACTGGCTTCTGCTGTGTAACAAGACCATCATCAATTTGACATCA 1490
657 GluIleGluHisArgIleAsnIleGluLysLeuLysAspAsn-LeuGlyIleHisTyrLys 676
1491 AGAAAAGTCAAGAACCTG-----CTTTCCACATTGCG----- 1524
676 sGlnIleAspGlyLeuGlnAsnGluMetSerGlnLysIleGluThrMetGlnPheGln 696
1525 -----GAGATCTTGTTCGAAGAAATAATCAATGTTGATCTGAGTAGTACCGATATATA 1580
696 uLysAspAsnLeuIleThrLysGlnAsnGlnLeuIleLeuGluIleSerLysLeuLys 716
1581 CAATGAGGTCTCCATCAACACCTTTCTGAGCTCAAGGAAATCCANAAGCTTAAAT 1640
716 p-----LeuGlnGlnSerLeuValAsnSerLysSerGluGluMetThrLeuGlnIle 733
1641 TAAT-----CTCAATATGACGAGATGCTCTAAGAGAAATAATACATGTTTCAGGACATG 1697
733 eAsnGluLeuGlnLysGluIleGluIleLeuArgGlnGlnGluLysGluLysGlyThr-L 753
1698 CACAAGAGACCAACCTGAAACACAGT-----CAATGAAG 1736
753 euGlnGlnGluValGlnGluLeuGlnLeuLysThrGluLeuGluLysGlnMetLysG 773
1737 AGCTGAACACATGTTTCAANCGAA----- 1762
773 lylsGluAsnAspLeuGlnGluLysPheAlaGlnLeuGluAlaGluAsnSerIleLeuL 793

```
QY 1763 -----CAAGATNATGTGAACAAACACACT-----GANCAGCAGG 1796
Db 793 ysAspGluLysThrLeuGluAspMetLeuLysIleHisThrProValSerGlnGluG 813
QY 1797 AGTCTCTTA-----GATCAGA 1811
Db 813 luArgLeuIlePheLeuAspSerIleLysSerLysSerLysAspSerValTrpGluLysG 833
QY 1812 AATTAATTTCACTACAAGCAAAATATGTGGCTTCAACAGCAATTAAGTTCATGCACATA 1871
Db 833 luileGluLeuLeuLeuGluAsnGluAspLeuLysGlnGlnCysIleGlnLeuAsnG 853
QY 1872 ANGAAGCTGCAACAAAGCAAGATAAATTCATNTTCATTTCTTGAGAGAAATG 1931
Db 853 luGluileGluLysGlnGlnAsnThrPheSerPheAlaGluLysAsnPhedluValAsnT 873
QY 1932 CNCATCATCTCTTAAAGAG----- 1951
Db 873 yrGlnGluLeuGlnGlnGluGluTyraLacCysLeuLeuLysValLysAspLeuGluAsps 893
QY 1952 --AAAAATGAGGAGATATTNTATTACNATAACCATTTAAAA 1990
Db 893 erLysAsnLysGlnGluLeuGluTyLysSerLysLeuLys 906

RESULT 6
GOA4_HUMAN
ID GOA4_HUMAN STANDARD; PRT; 2230 AA.
AC Q13439; Q13270; Q13654; Q14436;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Golgi autoantigen, golgin subfamily A member 4 (Trans-Golgi p230) (256
DE kDa golgin) (Golgin-245) (72.1 protein).
GN GOLGA4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=96215236; PubMed=8626529;
RA Erlich R., Gleeson P.A., Campbell P., Dietzsch B., Toh B.-H.;
RT "Molecular characterization of trans-Golgi p230: a human peripheral
RT membrane protein encoded by a gene on chromosome 6p12-22 contains
RT extensive coiled-coil alpha-helical domains and a granin motif.";
RL J. Biol. Chem. 271:8328-8337(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Seelig H.P.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBSJ databases.
RN [3]
RP SEQUENCE OF 131-2230 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=96125112; PubMed=8537393;
RA Fritzler M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.L.;
RT "Molecular characterization of golgin-245, a novel Golgi complex
RT protein containing a granin signature.";
RL J. Biol. Chem. 270:31262-31268(1995).
RN [4]
RP SEQUENCE OF 524-672 FROM N.A.
RC TISSUE=Gastric fundus;
RA Balaque C.;
RL Thesis (1994), Instituto municipal de investigacion medica, Spain.
CC -!- FUNCTION: May play a role in vesicular transport from the trans-
CC Golgi.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic, peripheral membrane protein
CC associated with the trans-Golgi network.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC isoId=Q13439-1; Sequence=Displayed;
CC
```

```
CC Name=2;
CC isoId=Q13439-2; Sequence=VSP_004272; VSP_004273;
CC Name=3;
CC isoId=Q13439-3; Sequence=VSP_004274;
CC Name=4;
CC isoId=Q13439-4; Sequence=VSP_004275;
CC -!- DOMAIN: Extended rod-like protein with coiled-coil domains.
CC -!- DISEASE: Antibodies against GOLGA4 are present in sera from
CC patients with Sjogren's syndrome (SS) [MIM:270150]. Sera from
CC patients with SS often contain antibodies that react with normal
CC components of the Golgi complex.
CC -!- DISEASE: Antibodies against GOLGA4 are found in sera from
CC hepatitis B patients.
CC -!- SIMILARITY: Belongs to the golgin family.
CC -!- SIMILARITY: Contains 1 GRIP domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U41740; AAC50434.1; -
CC EMBL; X82834; CAAS8041.1; -
CC EMBL; U31906; CAAS1791.1; -
CC EMBL; X76942; CAAS4261.1; -
CC Genew; HGNC:4427; GOLGA4.
CC MIM; 602509; -
CC MIM; 270150; -
CC GO; GO:0005802; C:Golgi trans face; TAS.
CC GO; GO:0016192; P:vesicle-mediated transport; TAS.
CC InterPro; IPR000237; GRIP_domain.
CC Pfam; PF01465; GRIP; 1.
CC PROSITE; PSS0913; GRIP; 1.
CC KW Antigen; Golgi stack; Coiled coil; Alternative splicing.
CC FT DOMAIN 133 2185
CC FT COILED COIL (POTENTIAL).
CC FT GLU-RICH.
CC FT GRIP.
CC FT TPYKGNLHYTDVSLFGPEPTFEYLRKVLPEY -> HLTKV
CC AICTRMHSLENLPNLSICEKFLSI (in isoform
CC 2).
CC /FTid=VSP_004272.
CC Missing (in isoform 2).
CC /FTid=VSP_004273.
CC Missing (in isoform 3).
CC /FTid=VSP_004274.
CC FTSPRGIF -> SWLRSS (in isoform 4).
CC /FTid=VSP_004275.
CC FTid=VSP_004275.
CC R -> K (IN REF. 3).
CC Y -> H (IN REF. 3).
CC T -> A (IN REF. 3).
CC K -> E (IN REF. 3).
CC T -> A (IN REF. 3).
CC K -> E (IN REF. 3).
CC K -> N (IN REF. 3).
CC SEQUENCE 2230 AA; 261139 MW; 3BB733DB1EA86134 CRC64;

Alignment Scores:
Pred. No.: 1.26e-07 Length: 2230
Score: 271.50 Matches: 159
Percent Similarity: 41.00% Conservative: 160
Best Local Similarity: 20.44% Mismatches: 286
Query Match: 7.98% Indels: 173
DB: 1 Gaps: 32

US-09-602-362E-15 (1-2030) x GOA4_HUMAN (1-2230)
QY 84 ATGCAAACTTTCAAGCAGAGCCCTCCGAGAGCCATCTCCCTCGAGCCCTGCATTGAA 143
Db 150 LeuSerSerTyArgGlyLysTySerSerGluLeuValThrAlaTyGln---MetLeuGln 168
```


Db 854 hrGlueAsAlaHisIysIleGlnValGlnAspLeuMetGlnGlnLeuGluLysGlnA 874

QY 1956 ATGACGAGATATTNATTACNATACCATTTTAAACCCCATATTTC 2005
|||::|
|||::|

Db 874 snSerGluMetGluGlnLysValLysSerLeuThrGlnValLysGlnSer 890
|||::|
|||::|

RESULT 7

TRIA_HUMAN STANDARD; PRT; 1979 AA.

ID TRIA_HUMAN
AC 015643; Q14689; O15154; O95949;
DT 01-NOV-1997 (Rel. 35, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Thyroid receptor interacting protein 11 (TRIP-11) (Golgi-associated
microtubule-binding protein 210) (GMAP-210) (Trip230) (Clonal
evolution related gene on chromosome 14).

GN TRIP11 OR CEV14.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND INTERACTION WITH RBL AND THRB.
RC TISSUE=Fibroblast;
RX MEDLINE=97404344; PubMed=9256431;
RA Chang K.-H., Chen Y., Chen T.-T., Chou W.-H., Chen P.-L., Ma Y.-T.,
Yang-Feng T.L., Leng L., Tsai M.-J., O'Malley B.W., Lee W.-H.;
RT "A thyroid hormone receptor coactivator negatively regulated by the
retinoblastoma protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:9040-9045 (1997).

RN [2]
RP SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, SUBCELLULAR
LOCATION, AND INTERACTION WITH MICROTUBULES.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=9207053; PubMed=10189370;
RA Infante C., Ramos-Morales F., Pediani F., Bornens M., Rios R.M.;
RT "GMAP-210, a cis-Golgi network-associated protein, is a minus end
microtubule-binding protein.";
RL J. Cell Biol. 145:83-98 (1999).

RN [3]
RP SEQUENCE FROM N.A.
RX PubMed=12508121;
RA Heilig R., Eckenberg R., Petit J.-L., Forknachten N., Da Silva C.,
Cattolico L., Levy M., Barbe V., De Berardinis V., Ureta-Vidal A.,
Pelletier E., Vico V., Anthouard V., Rowen L., Madan A., Qin S.,
Sun H., Du H., Pepin K., Attiguenave F., Robert C., Cruaud C.,
Bruls T., Jallou O., Friedlander L., Samson G., Brottier P.,
Cure S., Segurens B., Aniere F., Samain S., Crespeau H., Abbasi N.,
Aiach N., Boscus D., Dickhoff R., Dorx M., Dubois I., Friedman C.,
Gouyvenoux M., James R., Madan A., Mairéy-Estrada B., Mangenot S.,
Martins N., Menard M., Oztas S., Ratcliffe A., Shaffer T., Trask B.,
Vacherie B., Bellemere C., Belser C., Besnard-Gonnet M.,
Bartol-Mavel D., Boutard M., Briex-Silla S., Combette S.,
Dufosse-Laurent V., Fertou C., Lechaplais C., Louesse C., Muselet D.,
Magdelenat G., Pateau E., Petit E., Sirvain-Trukniewicz P., Trybou A.,
Vega-Czarny N., Bataille E., Bluet E., Bordelais I., Dubois M.,
Dumont C., Guerin T., Hafray G., Hammadi R., Munga J., Pellouin V.,
Robert D., Wunderle E., Gauguet G., Roy A., Sainte-Marthe L.,
Verdier J., Verdier-Discala C., Hillier L.W., Fulton L., McPherson J.,
Matsuda F., Wilson R., Scarpetti C., Gyapay G., Winkler P., Saurin W.,
Quetier F., Waterston R., Hood L., Weissbach J.;
RT "The DNA sequence and analysis of human chromosome 14.";
RL Nature 421:601-607 (2003).

RN [4]
RP SEQUENCE OF 1198-1979 FROM N.A., CHROMOSOMAL TRANSLOCATION WITH
PDGFRB, AND TISSUE SPECIFICITY.
RC TISSUE=Leukemia;
RX MEDLINE=98043615; PubMed=9373237;
RA Abe N., Emi N., Tanimoto M., Terasaki H., Marunouchi T., Saito H.;
RT "Fusion of the platelet-derived growth factor receptor beta to a novel
gene CEV14 in acute myelogenous leukemia after clonal evolution.";
RL Blood 90:4271-4277 (1997).

RN [5]

RP SEQUENCE OF 1757-1944 FROM N.A., AND INTERACTION WITH THRB.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=95295737; PubMed=7776974;
RA Lee J.W., Choi H.-S., Gyuris J., Brent R., Moore D.D.;
RT "Two classes of proteins dependent on either the presence or absence
of thyroid hormone for interaction with the thyroid hormone
receptor.";
RL Mol. Endocrinol. 9:243-254 (1995).

CC -!- FUNCTION: Binds the ligand binding domain of the thyroid receptor
(THRB) in the presence of triiodothyronine and enhances THRB-
modulated transcription. Golgi auto-antigen; probably involved in
maintaining cis-Golgi structure.

CC -!- SUBUNIT: Binds RBL.

CC -!- SUBCELLULAR LOCATION: Peripheral membrane protein associated with
the cis-Golgi network. Associates with the ends of centrosome-
nucleated microtubules.

CC -!- TISSUE SPECIFICITY: Highly expressed in pancreas, muscle, heart,
testis, peripheral blood leukocytes, and in several leukemia cell
lines. Detected at intermediate levels in placenta and kidney, and
at low levels in brain and lung.

CC -!- DOMAIN: Extended rod-like protein with coiled-coil domains
of leukemia and eosinophilia.

CC -!- DISSEASE: Involved in a acute myelogenous leukemia through a
chromosomal translocation t(5;14)(q33;q32) that involves PDGFRB
and TRIP11. The fusion protein may be involved in clonal evolution
of leukemia and eosinophilia.

CC -!- SIMILARITY: Belongs to the golgin family.

CC -!- SIMILARITY: Contains 1 GRIP domain.

CC -!- CAUTION: Ref.4 sequence differs from that shown due to
frameshifts in position 1932 and 1955.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

CC -----

DR EMBL; AF007217; AAD09135.1; -.
DR EMBL; Y12490; CAAV3095.1; -.
DR EMBL; AL049872; -; NOT ANNOTATED CDS.
DR EMBL; AF011368; AAB84386.1; ALT FRAME.
DR EMBL; L40380; AAC41730.1; ALT_SEQ.
DR PR; T03719; T03719.
DR HSP; P04266; IIC2.
DR Genew; HGNC:12305; TRIP11.
DR MIM; 604505; -.
DR GO; GO:0005515; F:protein binding; TAS.
DR GO; GO:0003713; F:transcription co-activator activity; TAS.
DR GO; GO:0006366; P:transcription from Pol II promoter; TAS.
DR InterPro; IPR003345; M_repeat.
DR Pfam; PF02370; M; 9.
DR PROSITE; PS50913; GRIP; 1.
DR DOMAIN; 52 1773
FT DOMAIN 1774 1823
FT SITE 1754 1755
FT CONFLICT 121 121
FT CONFLICT 382 382
FT CONFLICT 391 391
FT CONFLICT 516 516
FT CONFLICT 561 561
FT CONFLICT 1202 1202
FT CONFLICT 1237 1237
FT CONFLICT 1346 1347
FT CONFLICT 1658 1665
FT CONFLICT 1670 1670
FT CONFLICT 1827 1827
SQ SEQUENCE 1979 AA; 227638 MW; 83E4677D34D360D CRC64;

Alignment Scores: 1.73e-07 Length: 1979
Pred. No.:

| | | | |
|--|--------|---|------|
| Score: | 269.00 | Matches: | 160 |
| Percent Similarity: | 41.98% | Conservative: | 149 |
| Best Local Similarity: | 21.74% | Mismatches: | 265 |
| Query Match: | 7.90% | Indels: | 162 |
| DB: | 1 | Gaps: | 32 |
| US-09-602-362E-15 (1-2030) x TRIA_HUMAN (1-1979) | | | |
| QY | 63 | AAAGCCTTAGAATGTAGACATGCAAACTTTCARAGCAGAGCCTCCCGAGAGCCATCT | 122 |
| DB | 292 | LysThrIleGlnValLeuGlnIleGlu-----LysValGluSerThrLysLysMetGlu | 309 |
| QY | 123 | GCCTTCGAGCCTCCCATGAATGCAAAAGTCTGTCTCCAAATAAAGCCCTTGGAAATTGAAG | 182 |
| DB | 310 | GlnLeuGluAspLysIle-----LysAspIleAsnLysLysLeuSerSerAlaGlu | 326 |
| QY | 183 | AATGAACAACATG-----AGAGCAGATGAGATCTCCATCAGATCCAAACAA----- | 233 |
| DB | 327 | AsnAspArgAspIleLeuArgGluGlnGluGlnLeuAsnValGluLysArgGlnIle | 346 |
| QY | 234 | ---AAGCACTATCAAGAAAGTCTTCGGGATCTGAGAGTCTC----- | 272 |
| DB | 347 | MetGluGluCysGluAsnLeuLysLeuGluCysSerLysLeuGlnProSerAlaValLys | 366 |
| QY | 273 | ---TGTGAGACTGTTTCACAGAGAGTGTGTGTTTACCCAGGCTCACATCAAAAGAA | 329 |
| DB | 367 | GlnSerAspThrMetThrGluLysGluArgIleLeuAlaGlnSerAlaSerValGluGlu | 386 |
| QY | 330 | ATAGATAAAATAAATGAAATTTAGAGAG----- | 359 |
| DB | 387 | ValPheArgLeuGlnGlnAlaLeuSerAspAlaGluAsnGluIleValArgLeuSerSer | 406 |
| QY | 360 | ---TCTCCTCTAAT-----GATGGTGTTCCTGAAGGCTCCCTGCAGATGAAGTT | 407 |
| DB | 407 | LeuAsnGlnAspAsnSerLeuAlaGluAspAsnLeuLysLeuLysMetArgIleGluVal | 426 |
| QY | 408 | TCTATTCCAACTAAAGCCTTA-----GAATTGATGGACATGCAAACTTTC | 452 |
| DB | 427 | LeuGluLysGluSerLeuLeuSerGlnGluLysGluGluGlnMetSerLeuLeu | 446 |
| QY | 453 | AAAGCAGAGCCTCCCGAGAGCCATCTGCTTCGAGCGCTGCCATTCGAAATGCMAAAGTCT | 512 |
| DB | 447 | Lys-----LeuAsnAsnGluTyrgluValIleLysSer | 457 |
| QY | 513 | GTTCGAAATTAAGCCTTGAATTAAGATGAATGAACAAACATTGAGAGCAGATCAGATGTC | 572 |
| DB | 458 | ThrAlaThrArgAspIleSerLeuAspSerGlu-----LeuHisAspLeuArgLeu | 474 |
| QY | 573 | CCTTCAGATCAAAACAAAGAGGTTGAAGAAATTCCTGGGATTCGAGAGTCTCCGT | 632 |
| DB | 475 | AsnLeuGluAlaLysGluGlnGluLeuAsnGlnSerIleSerGluLysGluThrLeuIle | 494 |
| QY | 633 | GAGACTGTTTCACAGAGGATGTGTGPACCCCAAGGCTACA----- | 674 |
| DB | 495 | AlaGluIleGluLeuAspArgGlnAsnGlnGluAlaThrLysHisMetIleLeuIle | 514 |
| QY | 675 | -----CATCAAAAGAAATGGAATAAATGAGTGGAAATTTAGAAAT | 716 |
| DB | 515 | LysAspGlnLeuSerLysGlnGlnAsnGluGlyAspSerIleIleSerLysLeuLysGln | 534 |
| QY | 717 | TCAACTGAGCTATCAAAATCTTGGATACAGTTCATCTTGTGAAAGCAAGGAACTT | 776 |
| DB | 535 | AspLeuAsnAspGluLys-----LysArgValHisGlnLeu | 546 |
| QY | 777 | CAAAAGATCACTGTCAACAACGTACAGAAATAATGGAACAAATCAAAAGAACTTTGT | 836 |
| DB | 547 | GluAspAsp-----LysMetAsp-----IleThrLysGluLeuAsp | 558 |
| QY | 837 | GTACTGAAAGAAACTGTGAGAGCAAAAGAAATAAATCACTAGTAGAAGCAAA | 896 |
| DB | 559 | ValGlnLysGluLysLeuIleGlnSerGluValAlaLeuAsnAspLeuHisLeuThrLys | 578 |
| QY | 897 | GTTAAATGGGAACAAGAGCTCTGCAGTGTGAGATTGACTTTAAACCAAGAGAAGAG--- | 953 |
| DB | 579 | GlnLysLeuGluAspLysValGluAsnLeuValAspGlnLeuAsnLysSerGlnGluSer | 598 |
| QY | 954 | -----AAGAGAAGAAATGCCGATATATTAATGAAATAATAGG----- | 992 |
| DB | 599 | AsnValSerIleGlnLysGluAsnLeuGlu---LeuLysGluHisIleArgGlnAsnGlu | 617 |
| QY | 993 | GAACAATTAAGAAATCGAAGACGACGACGATAGGAAGAGTTAGAAGTGAACAAACAACCTT | 1052 |
| DB | 618 | GluGluLeuSerArgIleArgAsn-----GluLeuMetGlnSerLeu | 631 |
| QY | 1053 | GAACAG-----GCTCTCAGAATAACAAGATATAGAAATG | 1085 |
| DB | 632 | AsnGlnAspSerAsnSerAsnPheLysAspThrLeuLeuLysGluArgGluAlaGluVal | 651 |
| QY | 1086 | AAGAGTGTAGAAAGTAAATTTGAATCAGGTTTTCACATCATGAAATGAAATTTATCTC | 1145 |
| DB | 652 | ArgAsnLeuLysGlnAsnLeuSerGluLeuGlnGlnLeuAsnGluAsnLeuLysVal | 671 |
| QY | 1146 | TTACATCAAAATTCGATGTTGAAAGAGAAATTCCTCATGCTAAACTGGAATAGCCACA | 1205 |
| DB | 672 | AlaPheAsp-----ValLysMetGluAsnGluLysLeuValLeuAlaCysGluAsp | 688 |
| QY | 1206 | CTGMAACACCAATACACGAA-----AAGGAAAT | 1235 |
| DB | 689 | ValArgHisGlnLeuGluCysLeuAlaGlyAsnAsnGlnLeuSerLeuGluLysAsn | 708 |
| QY | 1236 | AAATACTTTCAGGACATTAAGATTTTAAAGAAAGAAATGCTGAACCTCAGATGACCTTA | 1295 |
| DB | 709 | ThrIleValGluThrLeuLysMetGluLys-----GlyGluIleGluAlaGluLeu | 725 |
| QY | 1296 | AAACTGAAGAGAAATCAATTAATAAAGGCACTCAATATATAGTGGCAGCTTAAAGTT | 1355 |
| DB | 726 | CysTrpAlaLysLysArgLeuLeuGluGluAlaAsnLysTyrgluLysThrIleGluGlu | 745 |
| QY | 1356 | CTGATAGCTGAGAACACAACTGCTCTCT----- | 1385 |
| DB | 746 | LeuSerAsnAlaArgAsnLeuAsnThrSerAlaLeuGlnLeuGluHisLeuIle | 765 |
| QY | 1386 | AAATTGAAGAAACAAAGCAAAAGAAATCACTAGAG-----GCAGAAAT | 1430 |
| DB | 766 | LysLeuAsnGlnLysLysAspMetGluIleAlaGluLeuLysLysAsnIleGluGlnMet | 785 |
| QY | 1431 | GAATCACACATCCTAGA-----CTGGCTTCTGCTCTACAAGACCATGATCA | 1478 |
| DB | 786 | AspThrAspHisLysGluThrLysAspValLeuSerSerSerLeuGluGlnLysGln | 805 |
| QY | 1479 | ATT-----GTCACATCAAGAAAGTCAAGAACCTGCTTTC | 1514 |
| DB | 806 | LeuThrGlnLeuIleAsnLysLysGluIle-PheIleGluLysLeuLysGluArgSerSe | 825 |
| QY | 1515 | CACATTGCAGGAGATGCTTGTTCGAAAGAAATAATGATGTTGATGTAGTAGTCCGAT | 1574 |
| DB | 825 | rLysLeuGlnGluLeuAspLysTyrgSerGln-----AlaLeu | 838 |
| QY | 1575 | ATATACAATGAGTGTCTCCATCAACCATTTCTGAAGCTCAAGGAAATCCANAAGCCT | 1634 |
| DB | 838 | uArgLysAsnGluIleLeuArgGlnThrIleGluGluLysAspArgSerLeuGlySerMe | 858 |
| QY | 1635 | AAAAATTAATCTCAATATTGAGGAGATGCTCTTAAGAGAAATACATTTGGTTTCAGGAAC | 1694 |
| DB | 858 | LysGluGluAsnAsnHisLeuGlnGluLeu-GluArgLeuArgGluGluGlnSerA | 878 |
| QY | 1695 | ATGCACAA-----AGAGACCAACGCGAAACACAGTGTCAAATGAAGGAGTGAACACA | 1748 |
| DB | 878 | rgThrAlaProValAlaAspProLysThrLeuAspSerValThrGluLeuAlaSerGluV | 898 |
| QY | 1749 | TGNTCAANCGAACAGATNATGTGAACAAACACACTGANCAGCAGGAGTCTTAGATC | 1808 |
| DB | 898 | alSerGlnLeu-----AsnThrIleLysGlnGluGluIleLysHisHisG | 916 |
| QY | 1809 | AGAAATTTTCAACTACAAAGCAAAATATGTGGCTTCAACAGCAATAGTTTCATGCAC | 1868 |

Db 916 InLysIleileGluAspGlnAsnGlnSer-----LysMetGlnLeuGlnSerL 933
Qy 1869 ATAANGAAGCTGACAAACAAAGCAAGATAACAATTCATNTTCATNTCTTGAGAGGAAA 1928
Db 933 euGlnGlu-----GlnLysLysGluMetAspGluPheArgTyrGlnHisGluGlnMetA 951
Qy 1929 ATGNCATCAT-----CTTCTAAAGAGAAAATGAGGAGATA 1966
Db 951 snAlaThrHisThrGlnLeuPheLeuGluLysAspGluGluile 965

RESULT 8
CENE_HUMAN STANDARD; PRT; 2663 AA.
ID CENE_HUMAN
AC Q02224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis.";
RL Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor.";
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
RN [4]
RP FARNESYLATION
RX MEDLINE=20459117; PubMed=10852915;
RA Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,
RA Bishop W.R., Kirschmeier P.;
RT "Farnesyl transferase inhibitors block the farnesylation of CENP-E
RT and CENP-F and alter the association of CENP-E with the
RT microtubules.";
RL J. Biol. Chem. 275:30451-30457(2000).
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -!- SIMILARITY: Belongs to the kinesin-like protein family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; Z15005; CAA78727.1; -
CC FIR; S28261; S28261.

HSP; P17119; 3KAR.
DR Genew; HGNC:1856; CENPE.
DR GK; Q02224; -
DR MIM; 117143; -
DR GO; GO:0005699; C:kinetochore; TAS.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0008350; F:kinetochore motor activity; TAS.
DR GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
DR GO; GO:0007079; P:mitotic chromosome movement; TAS.
DR GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN_MOTOR_DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere; Lipoprotein; Prenylation.
FT DOMAIN 1 335 KINESIN-MOTOR.
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP_BIND 86 93 ATP (BY SIMILARITY).
FT LIPID 2660 2660 S-farnesyl cysteine.
SQ SEQUENCE 2663 AA; 312087 MW; CSECI3880C8C8B8 CRC64;

Alignment Scores:
Pred. No.: 2,25e-07 Length: 2663
Score: 266.50 Matches: 172
Percent Similarity: 38.30% Conservative: 126
Best Local Similarity: 22.11% Mismatches: 305
Query Match: 7.83% Indels: 175
DB: 1 Gaps: 30

US-09-602-362E-15 (1-2030) x CENE_HUMAN (1-2663)
Qy 12 AAAGATGCTTCTCGAAGCTAACTCGGAATGAAGTTCTATTCCAACTAAGCCTTA 71
Db 1111 LysGluGlyGluLeuSerArgThrCysAspArgLeuAlaGluValGluGluLys 1130
Qy 72 GAATTGATGACATGCAAACTTTCAAAGCAGAGCCTCCGAGAGGCATCTGCCTTCGAG 131
Db 1131 Glu-----LysSerGlnGlnLeuGlnGlnGlnLeuLeuValGlnGlu 1148
Qy 132 CTGCGCATTAAGAAATGCAAAAG-----TCTGTTCCAAATAAAGCCTTGAATGAAG 182
Db 1149 GluMetSerGluMetGlnLysLysIleAsnGluIleGluAsnLeuLysAsnGluLys 1168
Qy 183 AAT---GAACAACATTGAGAGCAGATGAGATATCCCATCAGATCCCAACCAACAAAG--- 236
Db 1169 AsnLysGluLeuThrLeuGluHisMetGluThrGluArgLeuGluLeuAlaGlnLysLeu 1188
Qy 237 -----GACTATGAAGAAAGTCTCTGGGATTCGTGAGAGTCTCTGTGAGACTGTTTCACAG 290
Db 1189 AsnGluAsnTyrGluGlu-----ValLysSerIle-----ThrLysGluArg 1202
Qy 291 AAGATGTGTGTTTACCAGAGGTACACATCAAAAGAAATAGATAATAATGGAATA 350
Db 1203 LysValLeuLysGluLeuGlnLysSerPheGluThrGluArgAspHisLeuArgGlyTyr 1222
Qy 351 TTGAAGAGTCTCTGTGATATGATGTTTCTCGAAGCTCCCTGCAGATGAAAGTTTCT 410
Db 1223 IleArgGlu----- 1225
Qy 411 ATTCAACTAAGCCTTAGAATTGATGGACATGCACAACTTTCAAGCAGAGCCTCCCGAG 470
Db 1226 -----IleGluAlaThrGlyLeuGlnThrLysGluLeuLysIleAlaHisIleHis 1243
Qy 471 AAGCCATCTGCCTCGAGCGCTGCATTGAATCAAAAGTCTGTTCCAAATAAAGCCTTG 530
Db 1244 LeuLysGluHisGlnGluThrIleAspGluLeuArgSerValSerGluLysThrAla 1263
Qy 531 GAATTGAAGATGAACAAACATTG----- 554

Db 1264 GlnIleAsnThrGlnAspLeuGluLysSerHisThrLysLeuGlnGluIlePro 1283
 QY 555 -----AGAGCAGATCAGATGTTCCCTTCAGAAATCAAAACAAAGAGGTTGAAGAA 605
 Db 1284 ValLeuHisGluGlnGluLeuLeuProAsnValLysLysValSerGluThrGlnGlu 1303
 QY 606 AATCTTGGGATTCAGAGTCTCCGTGAGACGTGTTTCAGAGAGATGTGTGTACCC 665
 Db 1304 ThrMetAsnGluLeuGluLeuThrGluGlnSerThrThrLysAspSerThrThrLeu 1323
 QY 666 AAGGCTACACATCAAAAGAAATG-----GATAAATAAGTCGAAATTAAGAAGT 716
 Db 1324 AlaArgIleGluMetGluArgLeuArgLeuAsnGluLysPheGlnGluSerGlnGluGlu 1343
 QY 717 TCACTAGCTATCAAAATCTTTGATACAGTTCATTCTTGTGAAGACGACGAGGAACTT 776
 Db 1344 IleLysSerLeuThrLysGluArgAspAsnLeuLysThrIleLysGluAlaLeuGluVal 1363
 QY 777 CAA-----AAAGATCACTGTGAACACGTACAGGAAATGGAACAAATGAA 824
 Db 1364 LysHisAspGlnLeuLysGluHisIleArgGluThrLeuAlaLysIleGlnGluSerGln 1383
 QY 825 AAGAAG-----TTTGTGTACTGAAAAAGAACTGTCAAGACCAAAAGAAATAAAA 875
 Db 1384 SerLysGlnGluGlnSerLeuAsnMetLysGluLysAspAsnGluThrThrLysIleVal 1403
 QY 876 TCACGTTAGACACCAAAA-----GTTAATGGGAACA 911
 Db 1404 SerGluMetGlnPheLysProLysAspSerAlaLeuLeuArgIleGluIleGluMet 1423
 QY 912 GAGCTCTGAGTGTGAGTGTGATTTAAACCAAGAGAGAGAGAG----- 956
 Db 1424 LeuGlyLeuSerLysArgLeuGlnGluSerHisAspGluMetLysSerValAlaLysGlu 1443
 QY 957 -----AGAAGAAATGCCGATATATTAATGAA 983
 Db 1444 LysAspLeuGlnArgLeuGlnGluValLeuGlnSerGluSerAspGlnLeuLysGlu 1463
 QY 984 AAAAAATAGGGAAGATTTAGGAGA----- 1007
 Db 1464 AsnIleLysGluIleValAlaLysHisLeuGluThrGluGluLeuLysValAlaHis 1483
 QY 1008 -----ATCCAGACGACGAT-----AGGAAA 1028
 Db 1484 CysCysLeuLysGluGlnGluThrIleAsnGluLeuArgValAsnLeuSerGluLys 1503
 QY 1029 GAGTTAGAA-----GTGAACAACACTTGACAG-----GCTCTC 1064
 Db 1504 GluThrGluIleSerThrIleGlnLysGlnLeuGluAlaIleAsnAspLysLeuGlnAsn 1523
 QY 1065 AGAATACAGATATAGAAATGAAGAGTGTAGAAAATAATTTGAATTCAGTTCACACT 1124
 Db 1524 LysIleGlnGluIleLysGluLysGluGlnLeuAsnIleLysGlnIleSerGluVal 1543
 QY 1125 CATGAAATGAAATATCTCTTACATGAAATTCATGTAAGAAAGAAATGCCATG 1184
 Db 1544 GlnGluAsnValAsn-----GluLeuLysGln 1552
 QY 1185 CTAAACTGGAATAGCCACACTGAACACCAATACAGGAAAGAAATAAATACTTT 1244
 Db 1553 PheLysGluHisArgLysAlaLysAspSerAlaLeuGlnSerIleGluSerLysMetLeu 1572
 QY 1245 GAGGACATTAAGATTTAAAGAAAAGATGCTGAATTCAGATGACCTAAATGAA 1304
 Db 1573 GluLeuThrAsnArgLeuGlnGluSerGlnGluIleGlnIleMetIleLysGluLys 1592
 QY 1305 GAGGAATCATTAACATAAGGAGCATCTCAATATAGTGGCAGCTTAAGATTCATGAGCT 1364
 Db 1593 GluGlu-----MetLysArgValGlnGlnAlaLeuGlnIleGluArgAspGlnLeuLys 1610
 QY 1365 GAGAACACAATG---CTCACTTCTAAATGAAGGAAAACAGACAAAGAAATACTAGAG 1421
 Db 1611 GluAsnThrLysGluIleValAlaLysMetLysGluSerGlnGluLysGlu-Tyr----- 1628

QY 1422 GCAGAAATTAATCACACCATCTAGATGGCTTCTGCTGTACAAGACCATGATCAAAT 1481
 Db 1629 -GlnPheLeuLysMetThrAlaValAsn-----GluThrGlnGluLysMe 1643
 QY 1482 GTGACATCAAGAAAGAAACCTGCTTTCCACATTCAGAGAGATGCTTTTGGCAA 1541
 Db 1643 TCysGluIleGluHisLeuLysGluGlnPheGluThrGlnLysLeuAsnLeuGluAsnIle 1663
 QY 1542 AGAAATTAATGATGTGTAGTAGTACCGATATATAACAATGAGGTGCTCCATCAACC 1601
 Db 1663 eGlu-----ThrGluAsnIleArgLeuThrGlnIleHisGluAs 1677
 QY 1602 ACTTTCTGAAGCTCAAGGAAATCC-----ANAAAGCTCAAAAATAA 1643
 Db 1677 nLeuGluGluMetArgSerValThrLysGluArgAspAspLeuArgSerValGluGluTh 1697
 QY 1644 TCTCAATATTCAGAGAGATCTCTAAGAGAAATACATGCTTTCAGGACATGCAACA 1703
 Db 1697 rLeuLysValGluArgAspGlnLeuLysGluAsn-----LeuArgGluThrIleThr-A 1715
 QY 1704 GAGAC-----CAACGTGAACACAG---TGTCAAATGAAGAGAGCTGAACATGT 1751
 Db 1715 rAspLeuGluLysGlnGluLeuLysIleValHisMetHisLeuLysGluHisGlnG 1735
 QY 1752 NTCAAACGACAA-----CATNATGTGAACAAACACACTGANCAGC 1793
 Db 1735 luThrIleAspLysLeuArgGlyIleValSerGluLysThrAsnGluIleSerAsnMetG 1755
 QY 1794 AGGATCTCTAGATCAGAAATATTTCACTCAAAAGCAAAATATGCTGCTTCAACAGC 1853
 Db 1755 lnyAspLeuGluHisSerAsnAspAlaLeuLysAlaGlnAspLeuLysIleGlnGluG 1775
 QY 1854 AATTA-----GTTATGACATGAAGAGCTGACAAACAAGCAAGATAACAATTGAT 1907
 Db 1775 luLeuArgIleAlaHisMetHisLeuLysGluGlnGlnGluThrIleAsp----- 1791
 QY 1908 NTTCAATNTCTTGAGAGAAATGCNATCATCTTCTTAAAGAGAGAAATAGGAGATAT 1967
 Db 1792 -----LysLeuArgGlyIleValSerGluLysThrAspLysLeuS 1805
 QY 1968 TTNATTACNATAACCATTTAAAAAACCGTATATTTCAATATGGAAGAAA 2017
 Db 1805 erAsnMetGlnLysAspLeuGluAsnSerAsnAlaLysLeuGlnGluLys 1821

RESULT 9

ID NUFL1_YEAST STANDARD; PRT; 944 AA.
 AC P32380;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE NUFL1 protein (Spindle poly body spacer protein SPC110).
 GN NUFL1 OR SPC110 OR YDR356W OR D9476.3.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=8288C;
 RX MEDLINE=92176232; PubMed=1541631;
 RA Mirzayan C., Copeland C.S., Snyder M.;
 RT "The NUFL1 gene encodes an essential coiled-coil related protein that
 is a potential component of the yeast nucleoskeleton.";
 RL J. Cell Biol. 116:1319-1332(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94064779; PubMed=7503995;
 RA Kilmartin J.V., Dyos S.L., Kershaw D., Finch J.T.;
 RT "A spacer protein in the Saccharomyces cerevisiae spindle poly body
 whose transcript is cell cycle-regulated.";
 RL J. Cell Biol. 123:1175-1184(1993).

```

[3]
RN SEQUENCE FROM N.A.
RC STRAIN-S288c / A3972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favallo A., Fullerton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifkin L., Riles L., Taich A., Trevaaskis E., Vignati D.,
RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC !- FUNCTION: Essential component of the nucleoskeleton. Potential
CC role in crosslinking filaments or anchoring other molecules. It
CC is essential for growth.
CC !- SUBCELLULAR LOCATION: NUCLEAR; TIGHTLY ASSOCIATED WITH THE
CC NUCLEUS. IT IS PRESENT IN A GRANULAR PATTERN THAT EXCLUDES THE
CC NUCLEOLUS.
CC !- PTM: May be regulated by phosphorylation.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z11592; CAAT7668.1; -.
DR EMBL; X73297; CAAS1733.1; -.
DR EMBL; U28372; AAB64791.1; -.
DR PIR; S26710; S26710.
DR GerMOnline; 140848; -.
DR SGD; S0002764; NUFL1.
DR GO; GO:0005823; C:central plaque of spindle pole body; IDA.
DR GO; GO:0005822; C:inner plaque of spindle pole body; IDA.
DR GO; GO:0005200; F:structural constituent of cytoskeleton; IPI.
DR GO; GO:0007020; F:microtubule nucleation; IPI.
KW Coiled coil; Nuclear protein; Phosphorylation.
FT DOMAIN 164 791
FT COILED COIL.
FT DOMAIN 54 59
FT NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 726 731
FT NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 742 747
FT NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 731 944
FT ARG/TYR-RICH.
SQ SEQUENCE 944 AA; 111781 MW; 04FAA074B8A0BC8 CRC64;

Alignment Scores:
Pred. No.: 2,738-07 Length: 944
Score: 266.00 Matches: 152
Percent Similarity: 41.73% Conservative: 138
Best Local Similarity: 21.87% Mismatches: 225
Query Match: 7.82% Indels: 180
DB: 1 Gaps: 28

US-09-602-362E-15 (1-2030) x NUFL1_YEAST (1-944)
QY 150 AAGTCGTGTTCCAAATAAAGCCCTGGAATGAAGATGAACAAACATTGAGAGCAGATGAG 209
Db LysThrValLysAspGlnValLeuGluLeuGluAsnSerAspValGlnSer----- 261
QY 210 ATACTCCCATCAGATCCCAACAAAGGACATGAAAGATTCCTGGGATTCAGAGT 269
Db -----LeuLysLeuArgSerLysGluAspGluLeuLysAsnLeuMetAsnGluLeuAsnGlu 280
QY 270 CTCTGTGAGACTGTTTCACAGAAGGATGTGTGTTT-----CCCAAGGCTACACATCAA 323
Db LeuLysSerAsnAlaGluGluLysAspThrGlnLeuGluPheLysLysAsnGluLeuArg 300
QY 324 AAAGAAATAGATAAATAAATGAATAGAGAGTCTCTGATAATGATGCTTTCTG 383
Db LysA-gThrAsnGluLeuAsnGluLeuLysLysLysLysLysLysLysLysLysLysLys 320
QY 384 AAGGCTCCCTGCAGATGAAGTCTTATTCACATGAAGCCTTA-----GAATTGATG 437

```

```

321 Lys-----GlnLysGlnAsnGluSerLysArgLeuLysAspGluLeuAsn 335
438 GACATGCAAACTTTCAAAGCAGAGCCCTCCGAGAGCCATCTGCCTTCGACCTGCACAT 497
336 GluLeuGluThrLysPheSerGluAsnGlySerGlnSerSerAlaLysGluAsnGluLeu 355
498 GAATATGCAAAAGCTCTGTTCAATTAAGCCCTGGAATTTGAAGATGAACAAACATTGAGA 557
356 LysMetLeuLys-----AsnLysLysLysLysLysLysLysLysLysLysLysLys 372
558 GCAGATCAGATGTTCCCTTCAGAAATCAAAACAAAGAGGTTTGAAGAAAATTTCTGGGAT 617
373 AsnSerGlnLeuLeuLeuLysGluLysLeuAlaSerLeuMetAlaGlnLeuThrGln 392
618 TCTGAGAGTCTCCGTGAGACTGTTTCAAGAGGATGTGTGTATCCCAAGGCTACACAT 677
393 LeuGluSer-----LysLeuAsnGlnArgAspSerGlnLeu-----GlySerArg 407
678 CAAAAGAAATGATGATAAAGTGAATAAAGTGAATAAAGTGAATAAAGTGAATAAAGTGA 737
408 GluGluLeuLeuLysLysThrAsnAspLysLeuGlnLysAspLysLysLysLysLys 426
738 TTGGATACAGTTCATCTTTGTGAAGAGCAAGGGAACCTTCAAAAGAGTCACTGTGAACAA 797
427 GluGluThrValSerLysAspGluArgLysLysLysLysLysLysLysLysLysLys 441
798 CTACAGGAAATGGAACAAATGAAAGAGAGTGTGTGTACTGAAAGAAAGTGAACCTGCA 857
442 -----LysValLysGlnLeuGluAsnAspLeuPheValLysLysLysLysLys 458
858 GAAGCAAAA-----GAATAAATATCAG-----TTA 884
459 GluSerLysThrIleThrAspAsnGluLeuGluSerLysAspLysLysLysLysLys 478
885 GAGAACCAA-----AAAGTT-----AAATGGGAACAGAGCTCTGCAGT 923
479 GluAsnAspLysValAlaGlnGluLysTyrSerLysMetLysLysLysLysLys 496
924 GTGAGATTGACTTTAAACCAAGAGAGAGAGAAATGCCGATATATTAATGAA 983
497 -----LysGluArgGluPheAsnTyrLysLysLysLysLysLysLysLys 508
984 AAATATTAGGAAGAAATTAGGAAGAAATCGAAGACGACATAGGAAGAGTTAGAGTGAAA 1043
509 LysLeuGluAspGluLysThrThrLeuAsnGluLysLysLysLysLysLysLysLys 528
1044 CAACAACTTCAACAGGCTCTCAGATACAGATATAGAAATGAAGAGTGTAGAAAGTAA 1103
529 SerGlnLeuLysAsn-----LysLysLysLysLysLysLysLysLysLysLys 537
1104 TTGAATCAGGTTTCTCAGACTCATGAAATGAAATATATCTTTACATGAAATTCATG 1163
538 ---AsnSerThrAlaThrHisHisMetLysGluAsnTyr----- 549
1164 TTGAAAAGAAATTCGCTGCTAAACCTGAATAGCCACACTGAACCAATACCAG 1223
550 ---GluLysGlnLeuGluSerLeuArgLysAspLysLysLysLysLysLysLysLys 568
1224 GAAAGGAAATTAATCTTTCAG----- 1247
569 AspSerGluAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 588
1248 -----GACATTAGATTTTAAAGAAAGAAATGCTGAACTTCAGATG 1289
589 SerGluLysArgSerLysAspLysLysLysLysLysLysLysLysLysLysLysLys 608
1290 ACCCTAAACTGAAGAGGAAATCATTAATCAAAAGGCGATCT----- 1331
609 AsnLeuLysLeuGlnGluAspGluLysSerSerLysLysLysLysLysLysLysLys 628
1332 -----CAATATAGTGGCGAGCTTAA 1352
629 LysAspPheAsnGlnLeuLysSerGluGlnSerAsnLysLysLysLysLysLysLysLys 648

```


between the actin and microtubule cytoskeleton. May link endocytic vesicles to microtubules. May play a role in formation of furrows during cellularization.

!- SUBUNIT: Interacts with Lva.

!- SUBCELLULAR LOCATION: Cytoplasmic; microtubule-associated. Lva-CLIP-190 complexes are found at the Golgi.

!- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=3;

Name=A;

isoId=Q9VJ5-1; Sequences=Displayed;

Name=B;

isoId=Q9VJ5-2; Sequences=VSP_050479;

Note=No experimental confirmation available;

Name=C;

isoId=Q9VJ5-3; Sequence=VSP_050480;

Note=No experimental confirmation available;

!- TISSUE SPECIFICITY: Specifically expressed at the tip of the furrow in cellularizing blastoderms. CLIP-190 and jar are coexpressed at several times in development and in a number of tissues, including embryonic axonal neuron processes and posterior pole.

!- SIMILARITY: Contains 2 CAP-Gly domains.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

EMBL; AF041382; AAB96783.1; -

EMBL; AE003655; AAF53604.1; -

EMBL; AE003655; AAF53605.2; -

EMBL; AE003655; AAN10987.1; -

EMBL; AF118896; AAM50756.1; -

EMBL; F890020503; CLIP-190.

GO; GO:0005938; C:cell cortex; IDA.

GO; GO:0005794; C:Golgi apparatus; IDA.

GO; GO:0005875; C:microtubule associated complex; IDA.

GO; GO:0003779; F:actin binding; IDA.

GO; GO:0008017; F:microtubule binding; IDA.

GO; GO:0005515; F:protein binding; IMP.

GO; GO:0007349; P:cellularization; IMP.

InterPro; IPR000938; CAP-Gly.

Pfam; PF01302; CAP_GLY; 2.

PROSITE; PS00845; CAP_GLY_1; FALSE_NEG.

PROSITE; PS0245; CAP_GLY_2; 2.

Cytoskeleton; Golgi stack; Microtubule; Actin-binding; Coiled coil; Repeat; Alternative splicing.

DOMAIN 143 185 CAP-GLY 1.

DOMAIN 260 302 CAP-GLY 2.

DOMAIN 378 468 COILED COIL (POTENTIAL).

DOMAIN 484 660 COILED COIL (POTENTIAL).

DOMAIN 488 1452 GLU-RICH.

DOMAIN 667 946 COILED COIL (POTENTIAL).

DOMAIN 926 981 COILED COIL (POTENTIAL).

DOMAIN 1001 1121 COILED COIL (POTENTIAL).

DOMAIN 1158 1549 COILED COIL (POTENTIAL).

DOMAIN 1200 1461 GLN-RICH.

DOMAIN 1565 1600 COILED COIL (POTENTIAL).

DOMAIN 348 348 Missing (in isoform B).

/FtId=VSP_050479.

MSDDTSAGGTGAPFPSPVTPADPEPCATASKLPGRIRSNIP

TPATGSGIPQSKMKAPSSFGTSGVSKIGRPPCCNHTTPK

SGPPPEATSMGRSDNLSINSAYT -> MSRESDDNLS

SINSAYTDLYQETVRFTRSSLSPTPDWDRFSPARRSLKSE

AGRSYDYVLEATGRRRS (in isoform C).

/FtId=VSP_050480.

S -> N (IN REF. 1).

D -> G (IN REF. 1).

K -> Q (IN REF. 1).

FT CONFLICT 561 561 E -> A (IN REF. 1).

FT CONFLICT 614 614 T -> S (IN REF. 1).

FT CONFLICT 683 683 S -> I (IN REF. 1).

FT CONFLICT 692 692 N -> Q (IN REF. 1).

FT CONFLICT 717 717 M -> K (IN REF. 1).

FT CONFLICT 769 769 F -> L (IN REF. 1).

FT CONFLICT 787 787 Q -> E (IN REF. 1).

FT CONFLICT 881 881 Q -> E (IN REF. 1).

FT CONFLICT 907 907 HLL -> QLQ (IN REF. 1).

FT CONFLICT 920 920 G -> E (IN REF. 1).

FT CONFLICT 929 929 C -> Y (IN REF. 1).

SQ SEQUENCE 1690 AA; 189063 MW; D6F7916A9C532F16 CRC64;

Alignment Scores:

Pred. No.: 3 87e-07 Length: 1690

Score: 262.50 Matches: 170

Percent Similarity: 37.96% Conservative: 131

Best Local Similarity: 21.44% Mismatches: 263

Query Match: 7.71% Indels: 229

DB: 1 Gaps: 34

US-09-602-362E-15 (1-2030) x CL190_DROME (1-1690)

Qy 39 GGAATGAAGTTCTTATTCACAACTAAAGCTTAAAGTTAGATTGACATG-----CAAACT 92

Dy 330 GlyThrMetAsnSerIleAlaThrAlaThrSerArgMetArgMetAsnAlaGln 349

Qy 93 TTCAAGCAGACGCTCCGAGAACCATCTGCTCGAGCTGCCATTGAAATG-----146

Dy 350 ArgLysSerSerThrProValLysProIleLeuAlaThrProLysSerGlnPheSerMet 369

Qy 146 -----146

Dy 370 GlnAspLeuLeuArgGluLysGlnGlnHisValGluLysLeuMetValGluArgAspLeu 389

Qy 147 ---CAAAAGTCTGTTCCAAATAAAGCTTGAATG---AAGATGAACAAACATGAGA 200

Dy 390 AspArgGluAspAlaGlnAsnGlnAlaLeuGlnLysAsnLeuAsnGluLeuLys 409

Qy 201 GCA-----GATGAGATCTCCCTCAAGATCCCAACAAAGAGACTATGAA 245

Dy 410 AlaArgIleValGluLeuGluSerAlaLeuAspAsnGluArgLysLysThrGluGluLeu 429

Qy 246 GAAAGTTCTTGGGATCTGAGAGTCTCTGT-----GAGACTTTTCACAGAGAGATGTC 299

Dy 430 GlnCysSerIleAspGluAlaGlnPheCysGlyAspGluLeuAsnAlaGlnSerGlnVal 449

Qy 300 TGTTTACCCAGGCTACACATCAAAAGAAATAGATAAAATAAATGAAATTA-----353

Dy 450 -----TyrLysGluLysIleHisAspLeuGluSerLysIleThrLys 463

Qy 354 -----GAAGAGTCTCTGTATATGAT 374

Dy 464 LeuValSerAlaThrProSerLeuGlnSerIleLeuProProAspLeuProSerAspAsp 483

Qy 375 GGTCTTTCTGAAG-----GCTCCCTGCGAAGTAAAGTTCTTATTCACCTAAAGCC 425

Dy 484 GlyAlaLeuGlnGluIleAlaLysLeuGlnLysMetThrIleGlnLysGlu 503

Qy 426 TTGAATTTGATG-----437

Dy 504 ValGluSerArgIleAlaGluGlnLeuGluGluArgLeuArgGluAsnValLys 523

Qy 438 -----GACATGCAAACTTCAAGCAGAGCTCCCGAGAGCCATCTGCCTTC 485

Dy 524 TyrLeuAsnGluGlnIleAlaThrLeuGlnSerGlnLeuValSerLysAspGluAlaLeu 543

Qy 486 GAGCTGCCATTGAAATGCAAAAGTCTGTTCCCAAT-----AAAGCTTTGGAATTTGAAG 539

Dy 544 GluLysPheSerLeuSerGluCysGlyIleGluAsnLeuArgGluLeuGluLeuLeu 563

Qy 540 AATGAACAACATTTGAGAGAGATGATGTTCCCTTCAGATCAACAAACAAAGAGGTT 599

Db 564 LysGluGluAsnGluLysGlnAlaGlnGluAlaGlnPheThrArgLysLeuAla 563
 QY 600 GAAGAAAT-----TCTGGGATTC 620
 Db 584 GluLysSerValGluValLeuArgLeuSerSerGluLeuGlnAsnLeuLysAlaThrSer 603
 QY 621 GAGAGTCTCCGT---GAGACTGTTTACAGAGGATGTGTGTGTACCCAGGCTACACAT 677
 Db 604 AspSerLeuGluSerGluArgValAsnLysThrAspGluCysGluLeuLeuGlnThrGlu 623
 QY 678 -----CAAAAGCAATGATTAATAAGTGA 704
 Db 624 ValArgMetArgAspGluGlnIleArgGluLeuAsnGlnLeuAspGluValThrThr 643
 QY 705 AAATTAGAA-----GATTCAACTAGCTATCAAAATCTTGATACAGTTTAT 752
 Db 644 GlnLeuAsnValGlnLysAlaAspSerSerAlaLeuAspMetLeu----- 659
 QY 753 TCTGTGAAGACAGGGAACCTTCAAAAGATCACTGTGAACACGTCACAGGAAAAATG 812
 Db 660 -----ArgLeuGlnLysGluGlyThrGluGluLysSerThrLeuLeu 673
 QY 813 GAACAAATGAAGAAGTTCGTGACTCAAAAGCAAACTGTCAGACCAAGAAATA 872
 Db 674 GluLys-----ThrGluLysGluLeuValGlnSerLysGluGln 686
 QY 873 AAATCACTAGTTAGAGAACCAAAAGTAAATGGGAACAAGAGCTCTGAGATTG 932
 Db 687 AlaAlaLysThrLeuAsnAspLysGluGlnLeuGluLysGlnIleSerAspLeuLysGln 706
 QY 933 ACTTTAAACCAAGACAGAGAGAGAAATGCCGATATATTAATAAGAAAAATTAGG 992
 Db 707 LeuAlaGluGlnLysLeuValArgGluMetThrGluAsnAlaIleAsnGlnIleGln 726
 QY 993 GAAGAATTAGGAATTCGAAGAGCAG-----CATAGGAAGAGTTAGAA----- 1037
 Db 727 LeuGluLysGluSerIleGluGlnGlnLeuAlaLeuLysGlnAsnGluLeuGluAspPhe 746
 QY 1038 GTCAAAACAACACTTGAACAGGCTCTCAGAAATACAGATATAGAA----- 1082
 Db 747 GlnLysLysGlnSerGluValHisLeuGlnGluIleLysAlaGlnAsnThrGln 766
 QY 1083 -----TTGAAGAGTGTAGAAAGTAAATTTGAAT 1109
 Db 767 LysAspPheGluLeuValGluSerGlyGluSerLeuLysLysGlnGlnGlnLeuGlu 786
 QY 1110 CAGGTTTCTCACATCATGAAATGAAATATCTCTTACATGAAATTCATGTGAA 1169
 Db 787 GlnLysThrLeuGlyHisGluLysLeuGlnAlaLeuGluGlu-----LeuLys 803
 QY 1170 AAGGAA-----ATTGCCATGCTAAACTGGAATAGCCACACTGAACACCAA 1217
 Db 804 LysGluLysGluThrIleLysGluLysGluGlnGlnGlnLeuGlnSerLys 823
 QY 1218 TACAGGAAAGGAAAT-----AAATCTTTAGACACTT 1253
 Db 824 SerAlaGluSerGluSerAlaLeuLysValGlnValGlnLeuGlnLeuGlnGln 843
 QY 1236 -----AAATCTTTAGACACTT 1253
 Db 844 GlnAlaAlaSerGlyGluGluGlySerLysThrValAlaLysLeuHisaspGluIle 863
 QY 1254 AAGATTTTAAAGAAAGAAATGCTGAATTCAGATGACCTTAACTCAAGAGAAATCA 1313
 Db 864 SerGlnLysSerGlnAlaGluGluThrGlnSerGluLysSerThrGlnSerAsn 883
 QY 1314 TTAACATAAAGGCATCTCAAT-----AGTGGCAGCTTAAAGTTCTGATAGCT 1364
 Db 884 LeuGluAlaLysSerLysGlnLeuGluAlaAlaAsnGlySerLeuGlu----- 899
 QY 1365 GAGACACAAATGCTCACTTCTAAATGAAGGAAAAACAGACAAAGAAATACAGAGCA 1424
 Db 900 GluGluAlaLysLysSerGlyHisLeuLeuGluGlnIleThrLys-----LeuLysSer 917

QY 1425 GAAATTTGAATCACACCACCTCTAGACTCGCTTCTGCTGTCAAGACCATGATCAATTTGTG 1484
 Db 918 GluValGlyGluThrGlnAlaLeuSerSerCys-----HisThrAspValGlu 934
 QY 1485 ACATCAAGAAAAAGTCAAGAACCTGCTTTCCACATTCACAGGAGATGCTTTGTCAAAGA 1544
 Db 935 SerLysThrLysGlnLeuGlu-----AlaAlaAsnAlaAlaLeuGlu--- 948
 QY 1545 AAAATGATGTTGAT-----GTGAGTAGTACCGATATATACCAATGAGGT 1589
 Db 949 LysValAsnLysGluThrAlaGluSerArgAlaGluAlaSerAspLeu-GlnAspLysVa 968
 QY 1590 G-----CTCCATCAACACCTTTCTGAAGCTCAAGAGAAATCCANAAG 1631
 Db 968 LysGluIleThrAspThrLeuHisAlaGluLeuGlnAlaGluArgSerSerSerAl 988
 QY 1632 CCTAAAATTAATCTCAATTATCGAGAGATGCTCTTAAGAGAAAATACAT----- 1681
 Db 988 aLeuHisThrLysLeuSerLysPheSerAspGluIleAlaThrGly-HisLysGluLeu 1008
 QY 1682 -----TGCTTTCAAGAACATGCACAAAGACCAACACGCTGAACACACAGT 1724
 Db 1008 hrSerLysAlaAspAlaTrpSerGlnGluMetLeuGlnLysGluLys----- 1023
 QY 1725 GTCAATGAAGAGAGCTGAACACATGNTCAANCGAACCAAGATNATGTGAACAAACACA 1784
 Db 1024 --GluGlnGlnLeuArgGlnGlnLeuGlnAspSerGlnAspSerGlnThrLysLeu 1043
 QY 1785 CTGANCAAGCAGGAGCTCTAGATCAGAAATTTTCAACTACAAAGCAAAATATGTGGC 1844
 Db 1043 ysAlaGluGlyGluArg---LysGluLysSerPheGluGluSerIleLysAsn-----L 1060
 QY 1845 TTCACACGCAATAGTTTCATGCACATAAAGAAAGC 1879
 Db 1060 euGlnGluGluValThrLysAlaLysThrGluAsn 1071
 RESULT 11
 ID MYH9_RAT STANDARD; PRT; 1961 AA.
 AC Q62812;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,
 DE type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
 GN MYH9.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Choi O.H., Park C., Itoh K., Adelstein R.S., Beaven M.A.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 CC FUNCTION: Cellular myosin appears to play a role in cytokinesis,
 CC cell shape, and specialized functions such as secretion and
 CC capping.
 CC SUBUNIT: Myosin is a hexameric protein that consists of 2 heavy
 CC chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2
 CC regulatory light chain subunits (MLC-2).
 CC DOMAIN: The rodlike tail sequence is highly repetitive, showing
 CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
 CC characteristic for alpha-helical coiled coils.
 CC SIMILARITY: Contains 1 myosin-like globular head domain.
 CC SIMILARITY: Contains 1 IQ domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)


```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U31463; AAA74950.1; -.
DR HSSP; P10587; 1BR2.
DR InterPro; IPR00048; IQ region.
DR InterPro; IPR001609; Myosin head.
DR InterPro; IPR004009; Myosin N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin N; 1.
DR Pfam; PF01576; Myosin tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS0096; IQ; 1.
DR PROSITE; PS0096; IQ; 1.
KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Multigene family.
FT DOMAIN 1 778
FT DOMAIN 779 808
FT DOMAIN 841 1927
FT NP_BIND 174 181
FT DOMAIN 654 676
FT SEQUENCE 1961 AA; 226336 MW; 959876D9681FB19E CRC64;
SQ
Alignment Scores:
Pred. No.: 4, 03e-07 Length: 1961
Score: 262.00 Matches: 164
Percent Similarity: 38.44% Conservative: 137
Best Local Similarity: 20.95% Mismatches: 284
Query Match: 7.70% Indels: 198
DB: 1 Gaps: 30
US-09-602-362E-15 (1-2030) x MYH9_RAT (1-1961)
QY 141 GAAATGCAAACTCTGTTCCAAATAAAGCTTGGAAATG---AAGATGAACAACATG 197
D 1066 GluLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluLeu 1085
QY 198 ACAGCA-----GATGAGATACTCCCATCAGATCCAAACAAAGGAC 239
D 1086 GlnAlaAlaLeuAlaArgValGluGluGluAlaAlaGlnLysAsnMetalLeuLys 1105
QY 240 TATGAAGAAAGTTCTTGGGATCTCAGAGTCTCTGTGACATGTTTTCACAGAAGATGTG 299
D 1106 IleArgGluLeuGluThrGlnIleSerGluLeuGlnGluAspLeuGluSerGluArgAla 1125
QY 300 TCTTTACCCAGGCTACACATCAAAA-----GAATAGATAAATAAT 344
D 1126 CysArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuAlaLeuLys 1145
QY 345 GGAATAATTAGAGAGTCTCTGATAATGATGTTT-----CTGAAGCTCCCTGC 395
D 1146 ThrGluLeuGluAspThrLeuAspSerThrAlaAlaGlnGlnGluLeuArgSerLysArg 1165
QY 396 AGAATGAAGATTTCTATCCAACTAAGCCATTAGATTGATGGACATCAACATTTCAA 455
D 1166 GluGlnGluValSerIleLeuLysLysThrLeuGlu---AspGluAlaLysThrHisGlu 1184
QY 456 CGAGAGCTCCCGAAGGCCATCTGCTTCGAGCTGCGCATTTGAA----- 500
D 1185 AlaGlnIleGlnGluMetArgGlnLysHisSerGlnAlaValGluGluLeuAlaGluGln 1204
QY 501 -----ATGCAAAAGTCTGTTCCCAAT 521
D 1205 LeuGluGlnThrLysArgValLysAlaThrLeuGluLysAlaLysGlnThrLeuGluAsn 1224
QY 522 AAAGCTTGAATTGAAGATCAACAAACATTGAGCGCAGATCAGATGTTCCCTTCAGAA 581
D 1225 GluArgGlyGluLeuAlaAsnGluValLysAlaLeuLeuGlnGlyLysGlyAspSerGlu 1244
QY 582 TCAAAACAAAGAGGTTGAAGAAAATTTCTGG-----GATTCT 620
```

```
1245 HisLysArgLysLysValGluAlaGlnLeuGlnLeuValLysPheSerGluGly 1264
QY 621 GAGAGTCTCGTGAGACTGTTTCCAGAGAGGATGTGTGTACCCAGGCTACACATCAA 680
D 1265 GluArgValArgThrGluLeuAlaAspLys-----ValSerLysLeu-----Gln 1279
QY 681 AAAGAAATGATAAATAAGTGA----- 704
D 1280 ValGluLeuAspSerValThrGlyLeuLeuAsnGlnSerAspSerLysSerLysLeu 1299
QY 704 ----- 704
D 1300 ThrLysAspPheSerAlaLeuGluSerGlnLeuGlnAspThrGlnGluLeuGlnGlu 1319
QY 705 -----AAATTAGAGATTCAACTAGCCTATCAAAAATCTTGATACAGTTTCAT 752
D 1320 GluAsnArgGlnLysLeuSerLeuSerThrLysLeuLysGlnMetGluAspGluLysAsn 1339
QY 753 TCTTGTGAAGAGCAAGGAACTTCAAAAAGATCCTGTGAACACAGCTACAGGAAAA--- 809
D 1340 SerPheArgGluGlnLeuGluGluGluAlaLysArgAsnLeuGluLysGln 1359
QY 810 -----ATGGAACAAATCAAAAAGAGTTT----- 833
D 1360 IleAlaThrLeuHisAlaGlnValThrAspMetLysLysMetGluAspGlyValGly 1379
QY 834 TGTGTA-----CTGAAA 845
D 1380 CysLeuGluThrAlaGluGluAlaLysArgArgLeuGlnLysAspLeuGluGlyLeuSer 1399
QY 846 AAGAACTGTGAGAGCAAGAAATAAATCAATCAGTTAGAGAACCAAAAGTAAATGG 905
D 1400 GlnArgLeuGluGluLysValAlaAlaTyrAspLysLeuGluLysThrLysThrArgLeu 1419
QY 906 GAACAAGAGCTCTGCAGTGTGAGATTGACTTTTAAACCAAGAA----- 947
D 1420 GlnGlnGluLeuAspLeuValAspLeuValAspLeuValAspLeuValSerAsn 1439
QY 948 ---GAAGAGAGAGAGAAATGCGGAT-----ATATTAAATGAA 983
D 1440 LeuGluLysLysGlnLysLysPheAspGlnLeuAlaGluGluLysThrIleSerAla 1459
QY 984 AAAATTAGGAAGATTTAGGAAGATCGA-----CAGCAGCATAGGAAGAG 1031
D 1460 LysTyrAlaGluGluArgAspArgAlaGluAlaGluAlaArgGluLysGluThrLysAla 1479
QY 1032 TTGAAGTGAACAAACAACTTGAACAGGCTCTCAGAAATCAAGATATAGATTG----- 1085
D 1480 LeuSerLeuAlaArgAlaLeuGluGluAlaMetGlu---GlnLysAlaGluLeuGluArg 1498
QY 1086 -----AAGAGTGTAGAAAGTAATTTGAATCAGGTTTCTCAGACTCATGAAATGAAAT 1139
D 1499 LeuAsnLysGlnPheArgThrGluMetGluAspLeuMetSerSerLysAspValGly 1518
QY 1140 TATCTCTTACATGAAATTTGCATGTTGAAAAGGAAATGCCATGCTAAACCTGGAATA 1199
D 1519 LysSerValHisGlu-----LeuGluLysSerAsnArgAlaLeuGluGlnVal 1535
QY 1200 GCCACACTGAACACCAATACACAGGAAAAAGGAAAAATAATAC-----TTTGAGGACATT 1253
D 1536 GluGluMetLysThrGlnLeuGluGluGluLeuAspGluLeuGlnAlaThrGluAspAla 1555
QY 1254 AAGATTTTAAAGAAAGAAT-----GCTGAACCTTCAGATGACCCATAAAA 1298
D 1556 LysLeuArgLeuGluValValAsnLeuGlnAlaMetLysAlaGlnPheGluArgAspLeuGln 1575
QY 1299 CTGAAGAGGAATCAATTAACATAAAGGCGCATCTCAATATAGTGGCAGCTTAAAGTTCTG 1358
D 1576 GlyArgAspGluGlnSerGluGluLysLysGlnLeuValArgGlnValArgGluMet 1595
QY 1359 ATAGCTGAG-----AACACATGCTCCTCTTAATTTGAAG 1394
```

Db 1596 GluAlaGluLeuGluAspGluArgLysGlnArgSerIleAlaMetAlaAlaArgLysLys 1615
 Qy 1395 GAAAAACAGACAAAGAAATATAGAGGAGGAGAAATTAATATCACCATCTAGACTGGCT 1454
 Db 1616 LeuGluMetAspLeuLysAspLeuGluAlaHisIleAspThrAlaAsnLysAsnArgGlu 1635
 Qy 1455 TCTGCTGTACACACCATGATCAATTTGTACATCAAGAAAAAGTCAAGAACTGCTTTC 1514
 Db 1636 GluAlaIle-LysGlnLeuArgLysLeu-----GlnAlaGlnMetLysAspCysMetArg 1653
 Qy 1515 CACATTGACGAGATGCTGTTTTCAGAAAGAAATGAATGTTGATGTGATGATACCGAT 1574
 Db 1653 GAspValAspThrArgAlaSerArgGlu----- 1663
 Qy 1575 ATATAAATGAGTGTCTCATCAACCACTTCTTGAAGCTCAAGGAGATCCANAAAGCT 1634
 Db 1664 -----GluIleLeu-----AlaGlnAlaLysGluAsnGluLys 1674
 Qy 1635 AAAAATTAATCTCAATTTATGACGAGATGCTCTTAAGAGAAAAATACATTTGTTTCAGGAAC 1694
 Db 1674 sLysLeuLys-SerMetGluAlaGluMetIle-----GlnLeuGlnGluGlu 1689
 Qy 1695 AT-----GCACAAAGACCACTGAAACACAGTGTCAATGAAGGAGCTGAAC 1745
 Db 1689 lubeuAlaAlaAlaGluArgAlaLysArgGlnAlaGln-----GlnGluArgAspG 1706
 Qy 1746 ACATGTTTCAAACGAAACAGATNATGTGAACAAACACACACTGACAGGAGTCT- 1801
 Db 1706 lubeuAlaAspGluIleAlaAsnSerSerClyLysGlyAlaLeuAlaLeuGluLysA 1726
 Qy 1802 -----CTAGATCAGAAATATTTCATCACTACAAACGAAAC 1835
 Db 1726 rgArgLeuGluAlaLeuIleAlaLeuLeuGluGluGluGluGlnGlnGlnGlnGln 1746
 Qy 1836 ATATGCTGCTTCAACGAGTATGTTTCATGCAATAAAGAAAGCTGACACAAAGCAAG 1895
 Db 1746 hrgLuleuIleAsnAspArgLeuLysAlaAsnLeuGlnIleAspGlnIleAsnThrA 1766
 Qy 1896 ATAAACAT-----TGATTTTCATTTCTGAGAGGAAATGCAATCATCTCTTCAAAG 1949
 Db 1766 spLeuAsnLeuGluArgSerHisAlaGlnLysAsnGluAsnAlaArgGlnLeuGluA 1786
 Qy 1950 AGAAATGAGGAGATTTTATACATACATTTTAAACACCGTATATTTCAATAT 2009
 Db 1786 rgGlnAsnLysGluLeuLysAlaLysLeuGlnGluMetGluSerAlaValLysSerLysT 1806
 Qy 2010 GGAAG 2014
 Db 1806 yLys 1807

RESULT 12

MYHB_RABIT STANDARD; PRT; 1972 AA.
 AC P35748;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Myosin heavy chain, smooth muscle isoform (SMHHC).
 GN MYH11.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92073350; PubMed=1961735;
 RA Babij P., Kelly C., Periasamy M.;
 RT "Characterization of a mammalian smooth muscle myosin heavy-chain
 RT gene: complete nucleotide and protein coding sequence and analysis of
 RT the 5' end of the gene";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10676-10680(1991).
 CC !- FUNCTION: Muscle contraction.
 CC !- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2

CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
 CC and 2 regulatory light chain subunits (MLC-2).
 CC !- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC !- DOMAIN: The rodlike tail sequence is highly repetitive, showing
 CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
 CC characteristic for alpha-helical coiled coils.
 CC !- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
 CC meromyosin (LMW) and 1 heavy meromyosin (HMM). It can later be
 CC split further into 2 globular subfragments (S1) and 1 rod-shaped
 CC subfragment (S2).
 CC !- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC !- SIMILARITY: Contains 1 IQ domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; M77812; AAA31395.1; -.
 DR HSSP; P10587; 1BR2
 DR InterPro; IPR000048; IQ region.
 DR InterPro; IPR001609; myosin_head.
 DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR002017; Spectrin.
 DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR PRODOM; PD000355; myosin_head; 1.
 DR SMART; SMC0015; IQ; 1.
 DR SMART; SMC0242; MYSC; 1.
 DR PROSITE; PS00096; IQ; 1.
 DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW Calmodulin-binding; ATP-binding; Methylation; Multigene family.
 FT DOMAIN 1 784 MYOSIN HEAD-LIKE.
 FT DOMAIN 785 807 IQ.
 FT DOMAIN 844 1934 COILED COIL (POTENTIAL).
 FT DOMAIN 1935 1972 CARBOXYL-TERMINAL.
 FT NP_BIND 178 185 ATP (POTENTIAL).
 FT DOMAIN 661 683 ACTIN-BINDING (BY SIMILARITY).
 FT DOMAIN 763 777 ACTIN-BINDING (BY SIMILARITY).
 FT MOD_RES 129 129 METHYLATION (TRI-) (POTENTIAL).
 SQ SEQUENCE 1972 AA; 227318 MW; 2061A224288D6A4C CRC64;

Alignment Scores:

Pred. No.: 4,28e-07 Length: 1972
 Score: 261.50 Matches: 128
 Percent Similarity: 40.00% Conservative: 118
 Best Local Similarity: 20.81% Mismatches: 215
 Query Match: 7.68% Indels: 154
 DB: 1 Gaps: 18
 US-09-602-362E-15 (1-2030) x MYHB_RABIT (1-1972)

Qy 528 TTGAATTGAAGATGAACAA----- 548
 Db 827 LeuLysLeuArgAsnTrpGlnTrpTrpArgLeuPheThrLysValLysProLeuLeuGln 846
 Qy 549 ---ACATTGAGAGCAGATCATGATGTTCCCTTCAGAATCAAAACAAAGAGGTTGAGAA 605
 Db 847 ValThrArgGlnGluGluMetGlnAlaLysGluAspGluLeuGlnLysLysGlu 866
 Qy 606 AATTCTTGGGATCTGAGAGTCTCCGTGAGACTGTTTTCACAGAGGATGTGTGTACCC 665
 Db 867 ArgGlnGlnLysAlaGluSerGluLeuGlnGlnGlnLysHisThrGlnLeu--- 895
 Qy 666 AAGCTACACATCAAAAGAAATGGAATAAATAGTGGAAATATTAGAGATTCAACTAGC 725

886 Db -----SerGluGluLysAsnLeuLeuGlnGlnGluLeuGlnAlaGluThrGlu 901
726 QY CTA-----TCRAAAATCTTG 740
902 Db LeuTyAlaGluAlaGluMetArgValArgLeuAlaAlaLysLysGlnGluLeuGlu 921
741 QY GATACAGTTCATCTTGTGAAGACAGCAAGGAACTTCAAAAAGATCATCTGTGAACAACGT 800
922 Db GGuileLeuHisGluMetGluAlaArgLeuGluGluGluGluAspArgGlyGlnGlnLeu 941
801 QY ACAGGAAATATGAACAAATGAAGAAGATTGCTGACTGAAAGAAAGAACTGTGACAA 860
942 Db GlnAlaGluArgLysLysMetAlaGlnGlnMetLeuAspLeuGluGluGlnLeuGluGlu 961
861 QY GCAAGAGAAATATAATACAGTTA----- 884
962 Db GluGluAlaAlaArgGlnLysLeuGlnGluLysValThrAlaGluAlaLysLys 981
885 QY -----GAGAACCAAAAGTTAAATGGGAACA 911
982 Db LysLeuGluAspLysLeuValMetAspAspGlnAsnLysLeuSerLysLysArg 1001
912 QY GAGCTCTGAGTGTGAGATTGACT-----TTAAACCAAGAAAGAGAGAG 956
1002 Db LysLeuGluGluArgLysSerLeuThrThrAsnLeuAlaGluGluGluGluLys 1021
957 QY AGAAGAAATGCCATATATTAATGAATAATAGGAAGATAGGAAGATCGAAG 1016
1022 Db AlaLysAsnLeuThrLysLysLysLysHisGluSerMetLysLeuGluVal 1041
1017 QY CAGCATGAGGAAGAGTTAGAACTTGAACCAACCTTGAACGGCTCTCAGAAATACAAG 1076
1042 Db ArgLeuLysLysGluGluLysSerArgGlnGluLeuGluLysLeuLysMetAsp 1061
1077 QY ATAGAATTGAAGAGTGAAGAAGTAATTTGAATCAGGTTTTCACACTCATGAAATGAA 1136
1062 Db GlyGluAlaSerAsp----- 1066
1137 QY AATTATCTCTCATGAAATATTCATGTTGAAAGAAATGCGCATCTGAACCTGAA 1196
1067 Db -----LeuHisGlu-----GlnleAlaAspLeuGlnAlaGln 1077
1197 QY ATAGCCACATGAAACACCAATACCAAGGAAAGAAATATAATCTTTGAGGACATTAAG 1256
1078 Db IleAlaGluLeuLysMetGlnLeuAlaLysLysLysGluGluGluLeuAlaAlaLeuAla 1097
1257 QY ATTTTAAAGAAAGAACTGCACTTCACTGATGACCTTAA----- 1298
1098 Db ArgLeuGluAspGluThrSerGlnLysAsnAsnAlaLeuLysLysLysLeuArgGluLeuGlu 1117
1299 QY -----CTGAAAGAGGAATCATTAATCTAAAGGGCATCTCAATATAGTGGG 1343
1118 Db GlyHisLysSerAspLeuGlnGluAspLeuAspSerGluArgAlaAlaArgAsnLysAla 1137
1344 QY -----CAGCTTAAAGTTCTCATAGCTGAG----- 1367
1138 Db GluLysGlnLysArgAspLeuGluGluLeuGluAlaLeuLysThrGluLeuGluAsp 1157
1368 QY -----AACCAATGCTCACTCTTAATTTGAAGGAAACCAAGCAACAAATA--- 1415
1158 Db ThrLeuAspThrThrAlaThrGlnGlnGlnGluLeuArgLysGluGlnGluValThr 1177
1416 QY -----CTAGAGCAGAAATTTGAATCACACCATCTGAGCTGCTCTGCT 1460
1178 Db ValLeuLysLysAlaLeuAspGluGluThrArgSerHisGluAlaGlnValGlnGluMet 1197
1461 QY GTACAGACCATGATCAAAATTTGACATCAAGAAAGTCAAGAACCTGCTTCCACATT 1520
1198 Db ArgGlnLysHisThrGlnValGluGluLeuThrGluGlnGluGlnPheLysArg 1217
1521 QY GCAGGAGATGCTGTTTGAAGAAAGAAATCAATGTTGATGTCAGTAGTACCATATATAA 1580
1218 Db AlaLysAlaAsnLeuAspLysThrLysGlnThrLeuGluLysGluAsnAlaAspLeu-Al 1237

1581 QY CAATGAG-----GTGCTCCATCAACCACTTCTGAAGCTCAAGGAAATCCAAAGCCT 1634
1237 Db aGlyLeuLeuArgValLeuGlyGlnAlaLysGlnGluValGluHisLysLysLysLe 1257
1635 QY AAAAAATTAATCTC-----AATTATGAGGAGATGCTCTTAAGAGAAATACATT 1682
1257 Db uGluValGlnLeuGlnGlnLeuGlnSerLysCysSerAspGlyLysArgAlaGln 1277
1683 QY GTTTTCAGGAACATGTCACAAAGAGACCAACGCTGAAACACACAGTGT-----CAATGA 1733
1277 Db uLeuAsn-AspLysValHisLysLeuGlnAsnGluValGluSerValThrGlyMetLeu 1297
1734 QY AGGAAGCTGAA-----CACATGT 1751
1297 Db erGluAlaGluGlyLysAlaLysLeuAlaLysGluValAlaSerLeuGlySerGlnL 1317
1752 QY NTCAAAACGACCAAGATNTGTGAACAAACACACTGANCAGCAGGAGTCTCTAGATCAGA 1811
1317 Db euGlnAspThrGlnGlnLeuGlnGlnGlnThrArgGlnLysLeuAsnValSerThrL 1337
1812 QY AATTATTTCAACTACAAAGCAAAATATGTGGTTTCAACAGCAATTAGTTTCATGCACATA 1871
1337 Db ysLeuArgGlnLeuGluAspGluArgAsnSerLeuGlnGlnGlnLeu-----A 1353
1872 QY ANGAAGCTGACACAAACGACATCAACAATTCATTTTCATTTCTTGAGAGAAATG 1931
1353 Db spGluGluMetGluAlaLysGln---Asn-LeuGluArgHisLysSerThrLeuAsnile 1371
1932 QY CNCATCATCTTCTAAAAGAGAAATGAGGAGATATTTTATTTATACNATAAACCATTTAAAA 1991
1372 Db GlnLeuSerAspSerLysLysLysLeuGlnAspPheAlaSerThrVal----- 1387
1992 QY ACCGTTATTTTCATATGCAATATGCAAAAAAANAANAANAANA 2030
1388 Db ---GluSerLeuGluGluGlyLysLysArgPheGlnLys 1399

RESULT 13
BEAL_HUMAN
ID_BEAL_HUMAN STANDARD; PRT; 1411 AA.
AC Q15075; Q14221;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Early endosome antigen 1 (Endosome-associated protein p162).
GN BEAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=95286647; PubMed=7768953;
RA Mu F.-T., Callaghan J.M., Steele-Mortimer O., Stenmark H.,
RA Parton R.G., Campbell P.L., McCluskey J., Yeo J.-P., Tock E.P.C.,
RA Toh B.-H.;
RT "BEAL, an early endosome-associated protein. BEAL is a conserved
RT alpha-helical peripheral membrane protein flanked by cysteine
RT 'fingers' and contains a calmodulin-binding IQ motif.";
RL J. Biol. Chem. 270:13503-13511(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Seelig H.P.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP INTERACTION WITH RAB5A
RX MEDLINE=98361239; PubMed=9697774;
RA Simonsen A., Lippe R., Christoforidis S., Gaullier J.-M., Brech A.,
RA Callaghan J.M., Toh B.-H., Murphy C., Zerial M., Stenmark H.;
RT "BEAL links PI(3)K function to Rab5 regulation of endosome fusion.";
RL Nature 394:494-498(1998).
RN [4]

INTERACTION WITH RAB5A AND RAB5B.
RX MEDLINE=99421752; PubMed=10491193;
RA Callaghan J.M., Nixon S., Bucci C., Toh B.-H., Stenmark H.;
RT "Direct interaction of EEA1 with Rab5b.";
RL Eur. J. Biochem. 265:361-366(1999).
[5]
RX INTERACTION WITH STX6, AND SUBCELLULAR LOCATION.
RX MEDLINE=99436077; PubMed=10506127;
RA Simonsen A., Gaullier J.-M., D'Arrigo A., Stenmark H.;
RT "The Rab5 effector EEA1 interacts directly with syntaxin-6.";
RL J. Biol. Chem. 274:28857-28860(1999).
[6]
RX MUTAGENESIS OF ASP-1352; ASN-1357; 1367-VAL-THR-1368; ARG-1375 AND
ARG-1400, HOMODIMERIZATION, AND INTERACTION WITH PHOSPHATIDYLINOSITOL
3-PHOSPHATE.
RX MEDLINE=99322673; PubMed=10394369;
RA Kutateladze T.G., Ogburn K.D., Watson W.T., de Beer T., Emr S.D.,
Burd C.G., Overduin M.;
RT "Phosphatidylinositol 3-phosphate recognition by the FYVE domain.";
RL Mol. Cell 3:805-811(1999).
[7]
RX MUTAGENESIS OF TRP-1349; CYS-1358; PHE-1365; ARG-1370; ARG-1371;
HIS-1372; HIS-1373; CYS-1374; ARG-1375; CYS-1377; GLY-1378; CYS-1385;
ARG-1400 AND CYS-1405, SUBCELLULAR LOCATION, AND INTERACTION WITH
PHOSPHATIDYLINOSITOL 3-PHOSPHATE.
RX MEDLINE=20387352; PubMed=10807926;
RA Gaullier J.-M., Roemling E., Gillooly D.J., Stenmark H.;
RT "Interaction of the EEA1 FYVE finger with phosphatidylinositol
3-phosphate and early endosomes. Role of conserved residues.";
RL J. Biol. Chem. 275:24595-24600(2000).
[8]
RX INTERACTION WITH RAB22A.
RX MEDLINE=21859373; PubMed=11870209;
RA Kauppi M., Simonsen A., Bremnes B., Vieira A., Callaghan J.M.,
Stenmark H., Olkkonen V.M.;
RT "The small GTPase Rab22 interacts with EEA1 and controls endosomal
membrane trafficking.";
RL J. Cell Sci. 115:899-911(2002).
[9]
RX MUTAGENESIS OF GLU-39; PHE-41; ILE-42; PRO-44; MET-47 AND TYR-60,
HOMODIMERIZATION, AND INTERACTION WITH RAB5C.
RX MEDLINE=22499593; PubMed=12493736;
RA Merithew E., Stone C., Bathiraj S., Lambright D.G.;
RT "Determinants of Rab5 interaction with the N terminus of early
endosome antigen 1.";
RL J. Biol. Chem. 278:8494-8500(2003).
[10]
RX X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1289-1411 IN COMPLEX WITH
PHOSPHATIDYLINOSITOL 3-PHOSPHATE, AND HOMODIMERIZATION.
RX MEDLINE=21617582; PubMed=11741531;
RA Dumas J.J., Merithew E., Sudharshan E., Rajamani D., Hayes S.,
Lowe D., Corvera S., Lambright D.G.;
RT "Multivalent endosome targeting by homodimeric EEA1.";
RL Mol. Cell 8:947-958(2001).
[11]
RX STRUCTURE BY NMR OF 1346-1410 ALONE AND IN COMPLEX WITH
PHOSPHATIDYLINOSITOL 3-PHOSPHATE.
RX MEDLINE=21143489; PubMed=11230696;
RA Kutateladze T.G., Overduin M.;
RT "Structural mechanism of endosome docking by the FYVE domain.";
RL Science 291:1793-1796(2001).
CC -!- FUNCTION: binds phospholipid vesicles containing
phosphatidylinositol 3-phosphate and participates in endosomal
trafficking.
CC -!- SUBUNIT: Homodimer. Binds STX6. Binds RAB5A, RAB5B, RAB5C and
RAB22A that have been activated by GTP-binding.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and peripheral membrane protein
associated with early endosomes.
CC -!- DOMAIN: The FYVE-type zinc finger domain mediates interactions
with phosphatidylinositol 3-phosphate.
CC -!- DISEASE: Antibodies against EEA1 are found in sera from patients
with subacute cutaneous lupus erythematosus and other autoimmune
diseases.

CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
CC -!- SIMILARITY: Contains 1 FYVE-type zinc finger.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; L40157; AAA79121.1; -;
DR EMBL; X78998; CAA55632.1; -;
DR PIR; A57013; A57013;
DR PDB; 1HY1; 14-MAR-01.
DR PDB; 1HYJ; 14-MAR-01.
DR PDB; 1JOC; 28-DEC-01.
DR Genew; HGNC:3185; BEA1.
DR MIM; 605070; -;
DR GO; GO:0005737; C:cytoplasm; TAS.
DR GO; GO:0005769; C:early endosome; TAS.
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0005545; F:phosphatidylinositol binding; TAS.
DR GO; GO:0006899; P:nonspecific vesicle transport; TAS.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR000306; Znf_FYVE.
DR Pfam; PF01363; FYVE; 1.
DR Pfam; PF00096; zf-C2H2; 1.
DR SMART; SM00064; FYVE; 1.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00178; ZF_FYVE; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 1.
DR Antigen; Metal-binding; Zinc; Coiled coil; Zinc-finger; 3D-structure.
FT DOMAIN 74 1348
FT DOMAIN 397 758
FT DOMAIN 937 1032
FT DOMAIN 1093 1231
FT ZN_FING 41 64
FT ZN_FING 1352 1410
FT MUTAGEN 39 39
FT MUTAGEN 41 41
FT MUTAGEN 42 42
FT MUTAGEN 44 44
FT MUTAGEN 47 47
FT MUTAGEN 60 60
FT MUTAGEN 1349 1349
FT MUTAGEN 1352 1352
FT MUTAGEN 1357 1357
FT MUTAGEN 1358 1358
FT MUTAGEN 1365 1365
FT MUTAGEN 1367 1367
FT MUTAGEN 1370 1370
FT MUTAGEN 1371 1371
FT MUTAGEN 1372 1372
FT MUTAGEN 1373 1373
FT MUTAGEN 1374 1374
C->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
R->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
R->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
H->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
H->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
C->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
VT->EE,GG: ABOLISHES PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
R->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
R->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
H->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
H->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
C->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING AND ENDOSOMAL LOCATION.

Db 839 ThrLysValLysProLeuLeuGlnValThrArgGlnGluGluMetGlnAlaLysGlu 858
QY 477 TCTGCTTCGAGCTGCCATTGAAATGCAAAAGTCTGTTCACAAATAAAGCTTCGAAATTG 536
Db 859 GluGluMetGlnLysIleThrGluArgGlnGlnLysAlaGluThrGluLeuLysGluLeu 878
QY 537 AAGATGAACAAACATTCAGAGCAGATCAGATGTTCCCTTCAGATCAACAAACAAAGAG 596
Db 879 GluGlnLysHisThrGlnLeuAlaGluLysThrLeuLeuGlnGlnLysGlnAla 898
QY 597 GTTGAAGAAATCTTGGGATTCGAGAGTCTCCGTGAGACTGTTTCACAGAGGATGTG 656
Db 899 GluThrGluLeuTyraGluSerGluGluMetArg 910
QY 657 TGTGTACCAAGCTACATCAAAAGAAATGGATAAAATTAAGTGGAAATTTAGAAGAT 716
Db 911 ---ValArgLeuAlaAlaLysLysGlnGluLeuGluGluLeu 923
QY 717 TCAACTAGCTATCAAAATCTTGATACAGTTCATCTTCTGGAAGAGCAGGGAAGCTT 776
Db 924 -----LeuHisGluMetGluAlaArgLeuGluGlu 933
QY 777 CAATAAGATCACTGTGAACAAGCTACAGGAAAAATGGACAAATGGAAGAAAGCTTTTGT 836
Db 934 GluGluAspArgGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGlnMetLeu 953
QY 837 GTACTGAAAGAAATCTGACAGCAAAAGAAATPAAATCACAGTGA 884
Db 954 AspLeuGluGlnLeuGluGluGluAlaAlaArgGlnLysLeuGlnLeuLys 973
QY 885 -----CAG 887
Db 974 ValThrAlaGluAlaLysIleLysLysLeuGluAspAspIleLeuValMetAspGln 993
QY 888 AACCAAAAGTTAAATGGGAACAAGAGCTGTCAGCTGTGAGATTGACT 935
Db 994 AsnSerLysLeuSerLysGluArgLysLeuLeuGluGluArgValSerAspLeuThr 1013
QY 936 ---TTAAACCAAGAGAGAGAGAGAGAAATGCCGATATATTAATAAGAAAAATAGG 992
Db 1014 AsnLeuAlaGluGlnGluLysAlaLysAsnLeuThrLysLeuLysSerLysHisGlu 1033
QY 993 GAAGAATTAGGAAGAAATCGAAGAGCAGTAGGAAGAGTTAGAGTGAACAAACAACTT 1052
Db 1034 SerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGlnLeu 1053
QY 1053 GAACAGCTCTCAGATCAAGATATAGATTGAGAGGTGAGAAGTAATTTGAAATCAG 1112
Db 1054 GluLysLeuLysArgLysLeuGlu 1063
QY 1113 GTTCTTCACACTCATGAAATGAAATATTCTCTTACATGAAATTCATGTTGAAAAAG 1172
Db 1064 AlaSerAspPheHisGlu 1069
QY 1173 GAAATGGCATGCTAAATCTGAAATAGCCACACTGAAACACCAATACAGGAAAGAA 1232
Db 1070 GlnIleAlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGlu 1089
QY 1233 AATAAATCTTTGAGCAGATTAAATTTTAAAGAAAGAAATGCTGGAATTCAGATGACC 1292
Db 1090 GluGluLeuGlnAlaAlaLeuAlaArgLeuAspGluGluIleAlaGlnIleAsnAsnAla 1109
QY 1293 CTAATA-----CTGAAGAGGAATCAATTAAT 1319
Db 1110 LeuLysLysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuSer 1129
QY 1320 AAAAGGGCATCTCAATATAGTGGG-----CAGCTTAA 1352
Db 1130 GluArgAlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGlu 1149
QY 1353 GTTCTGATAGCTGAG-----AACCAATCTCCTCACTTCAAATTTGAAG 1394
Db 1150 AlaLeuLysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnLeuArg 1169

RESULT 15

USOL_YEAST STANDARD; PRT; 1790 AA.
ID USOL_YEAST
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Intracellular protein transport protein USOL.
GN USOL OR INTL OR YJL058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
CX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol1, is required for intracellular

QY 1395 GAAAAACAAGACAAAGAAATA-----CTAGAGCGAAGAAATGAATCA 1436
Db 1170 AlaLysArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluThrArgSer 1189
QY 1437 CACCATCTAGACTGCTTCTGTACAGACCATGATCAATTCGTGACATCAAGAAAA 1496
Db 1190 HisGluAlaGlnValGlnGluMetArgGlnLysHisThrGlnAlaValGluLeuThr 1209
QY 1497 AGTCAAGAACCTCTTTCACATTCGAGAGATGCTTGTTCACAAAGAAAAATGAATGT 1556
Db 1210 GluGlnLeuGlnPheLysArgAlaLysAlaAsnLeuAspLysSerLysGlnThrLeu 1229
QY 1557 GATGTGAGTAGTACCATATATAACAATGAG-----GTGCTCCATCAACCACTTCTGA 1610
Db 1230 GluLysGluAsnAlaAspLeu-AlaGlyGluLeuArgValLeuGlyGlnAlaLysGln 1249
QY 1611 AGCTCAAGAAATCCANAAGCCTAAAAATTAATCTC-----AATTATCAGG 1658
Db 1249 uValGluHisLysLysLysLysLeuGluValGlnLeuGlnAspLeuGlnSerLysCys 1269
QY 1659 AGATGCTCTAAGAGAAAAATACATTTGTTTCAGGACATGCACAAAGACCAACCTGAAA 1718
Db 1269 rAspGlyGluArgAlaArgAlaGluLeuSer-AspLysValHisLysLeuGlnAsnGlu 1289
QY 1719 CACAGTGT-----CAAAATGAAGGAAGCTGAA 1744
Db 1289 alGluSerValThrGlyMetLeuAsnGluAlaGluGlyLysAlaIleLysLeuAlaLys 1309
QY 1745 -----CACATGNTCAANCGAACCAAGATNATGTGAACAAACACACTG 1787
Db 1309 spValAlaSerLeuGlySerGlnLeuGlnAspThrGlnLeuLeuGlnGluGluThrA 1329
QY 1788 ANCAGCAGAGTCTTAGATCAGAAATTTTCAACTCAACAAACAAATATGTGCTTC 1847
Db 1329 rgGlnLysLeuAsnValSerThrLysLeuArgGlnGluAspGluArgAsnSerLeuG 1349
QY 1848 AACAGCAATTAGTTTCATGCACATAANGAAGCTGCACACAAACAAACAAAGATAACAATTGAT 1907
Db 1349 iAspGlnLeu-----AspGluGluMetGluAlaLysGln---Asn-LeuGlu 1363
QY 1908 NTTCATNTCTTGAGAGGAAATGCNCATCATCTTTAAAGAGAAAAATGAGAGATAT 1967
Db 1364 ArgHisValSerThrLeuAsnIleGlnLeuSerAspSerLysLysLysLeuGlnAspPhe 1383
QY 1968 TTNATTACNATAACCATTTAAAAAACCCTATATTTCAATATATGAAAAAANAANA 2027
Db 1384 AlaSerThrIle-----GluValMetGluGluGlyLysArgLeuGln 1398
QY 2028 AAA 2030
Db 1399 Lys 1399

RT protein transport in Saccharomyces cerevisiae.";

RL J. Cell Biol. 113:245-260(1991).

RN [2]

RP SEQUENCE OF 782-1790 FROM N.A.

RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,

RL Kendrick K.B.;

RA Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 1-8 FROM N.A.

RA Bai Y., Symington L.S.;

RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: Required for protein transport from the ER to the Golgi complex.

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE ER AND THE GOLGI COMPLEX.

CC -1- DOMAIN: The rodlike tail sequence is highly repetitive, composed of an heptapeptide repeat pattern characteristic of alpha-helical coiled coils. May form filamentous structures in the cell.

CC -1- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

DR EMBL; X54378; CAA38253.1; -

DR EMBL; L03188; AAB00143.1; -

DR EMBL; U53688; AAB66659.1; -

DR GenOnline; 140300; -

DR SGD; S0002216; USO1.

DR InterPro; IPR008938; ARM.

DR InterPro; IPR02017; Spectrin.

DR InterPro; IPR006955; USO1 p115 C.

DR InterPro; IPR006953; USO1 p115_head.

DR Pfam; PF04871; USO1_p115_C; 1.

DR Pfam; PF04869; USO1_p115_head; 1.

KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.

FT DOMAIN 1 724 GLOBULAR HEAD.

FT DOMAIN 725 1790 COILED COIL (POTENTIAL).

FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).

FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.

FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).

FT CONFLICT 847 847 G -> E (IN REF. 2).

FT CONFLICT 924 924 E -> K (IN REF. 2).

FT CONFLICT 1253 1253 V -> I (IN REF. 2).

FT CONFLICT 1319 1319 I -> V (IN REF. 2).

FT CONFLICT 1461 1461 N -> S (IN REF. 2).

FT CONFLICT 1581 1581 G -> S (IN REF. 2).

FT CONFLICT 1600 1600 I -> V (IN REF. 2).

FT CONFLICT 1661 1661 R -> S (IN REF. 2).

FT CONFLICT 1772 1772 D -> DEEDDE (IN REF. 2).

SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

Alignment Scores:

| | | | |
|------------------------|----------|---------------|------|
| Pred. No.: | 7,92e-07 | Length: | 1790 |
| Score: | 256.50 | Matches: | 163 |
| Percent Similarity: | 39.45% | Conservative: | 136 |
| Best Local Similarity: | 21.50% | Mismatches: | 298 |
| Query Match: | 7.54% | Indels: | 161 |
| DB: | 1 | Gaps: | 28 |

US-09-602-362E-15 (1-2030) x USO1_YEAST (1-1790)

QY 51 TCTATCCAACT-----AAAGCCCTAGATGATGAGATGCAACATTTCAAGACAG 104

DB 840 SerIleLysThrLeuGluLysGlyLeuGluThrIleLeuSerGlnLysLysLysAlaGlu 859

QY 105 CCT-----CCGAGAAGCCATCTGCCTTCGAGCGCTGCATTGAAATGCAAAAGTCT 155

860 AspGlyIleAsnLysMetGlyLysAspLeuPheAlaLeuSerArgGluMetGlnAlaVal 879

156 GTTCCAAATAAGCCTTGGATTGAAGAATGAACAAACATTGAGAGCAGATGAGATACTC 215

880 GluGluAsn----- 882

216 CCATCAGATCCAAACAAAGAGACTATGAAGAAGTTCTTGGGATCTGAGACTCTCTGT 275

883 ---CysLysAsnLeuGlnLysGluLysAspLysSerAsnValAsnHisGlnLys----- 899

276 GAGACTCTTTACAGAGAGGTGTGTTTACCAAGCTACACATCAAAAGAAATAGAT 335

900 GluThrLysSerLeuLysGluAspIleAlaLysIleThr-----GluIleLys 916

336 AAAATAATGAAAAATTAGAGAGTCTCTCGATAATGATGTTTCTGAAGGCTCCCTGC 395

917 AlaIleAsnGluAsnLeuGlu-----MetLysIleGlnCys 929

396 -----AGATGAAGTTTCTATTCCAACCTAAAGCCCTAGAAATGTGGACATG 443

930 AsnAsnLeuSerLysGluLysGluHisIleSerLysGluLeuValGlu----- 945

444 CAAACTTTCAAGCAGAGCCTCCGAGAGCCATCTGCCTTCGAGCCTGCCATTGAAATG 503

946 -----TyrLysSerArgPheGlnSerHisAspAsnLeuValAlaLysLeuThrGluLys 963

504 CAAAGCTCTTCCAAAT-----AAAGCTTGGAAATGAAGATCAACAAACATTGAGA 557

964 LeuLysSerLeuAlaAsnAsnTyrLysAspMetGlnAlaGluAsnGluSerLeuLys 983

558 GCAGATCAGATGTTCCCTTCAGAAATCAAAA-----CAAAAGAGGTT 599

984 AlaValGluLysLysAsnGluSerIleGlnLeuSerAsnLeuGlnAsnLysIle 1003

600 GAAGAAAAATTTCTGGGATCTTGAGAT----- 626

1004 AspSerMetSerGlnGluLysGluAsnPheGlnIleGluArgGlySerIleGluLysAsn 1023

627 -----CTCCGTGAGACTGTTTCACAG-----AAGGATGCTGTGTACCC 665

1024 IleGluGlnLeuLysLysThrIleSerAspLeuGluGlnThrLysGluLysIleSer 1043

666 AAGGCT-----ACACATCAAAAGAAATGATAAAATAAGTGAAATTA 710

1044 LysSerAspSerLysAspGluTyrGluSerGlnIleSerLeuLysGluLysLeu 1063

711 GAAGATCAACTAGC-----CTATCAAAATCTTGGATACA 746

1064 GluThrAlaThrThrAlaAsnAspGluAsnValAsnLysIleSerGluLeuThrLysThr 1083

747 GTTCATTCTTGTGAAAGAGCAGGGAACCTCAAAAAGATCACTGTGCAACAACTACAGGA 806

1084 ArgGluLysLeuGluAlaGluLeuAlaIleTyrLysAsnLeuLysAsnGluLeuThr 1103

807 AAAATGCAACAAATGAAAGAGTTTGTGACTGTAAGAAAGAACTGTCAAGAGCAAAA 866

1104 LysLeuGluThrSerGluLysAlaLeuLysGluValLysGluAsnGluLysIleLys 1123

867 GAAATAAAATCACAGTTTAGAACCAACAAAGTTTAAATGGGAACAGAGCTCTGCAGTGTG 926

1124 GluLysIleGlnLeuGluLysGluAlaThrGluThrLysGlnGlnLeuAsnSerLeu 1143

927 AGATTGACTTTAAACCAAGAGAGAGAGAGAGAAATGCCGATATATTAATGAAAAA 986

1144 ArgAlaAsnLeuGluSerLeuGluLysGluHisGluAspLeuAlaGlnLeuLysLys 1163

987 ATTAGGGAAGATTAGGAAGATCGAGAGCAGCATAGGAAGAGATTAGAA----- 1037

1164 TyrGluGluGlnIleAlaAsnLysGluArgGlnTyrAsnGluLysIleSerGlnLeuAsn 1183

1038 -----GTGAAACAAACAACTTGAACAGGCTCTCAGAAATCAAGATAGATTAATTTG 1085

1184 AspGluIleThrSerThrGlnGlnGluAsnGluSerIleLysLysLysAsnAspGluLeu 1203

Search completed: July 15, 2004, 09:08:50
Job time : 86.1613 secs

```
QY 1086 AAGAGTGTAGAAAGTAATTTGAATCAGAGTGTCTCAGACTCATGAAATGAAATATATCTC 1145
Db      |||      :|:      |||      |||      |||      |||      |||      |||
1204 -----GluGlyGluValLysAlaMetLysSerThrSerGluGluGlnSerAsnLeu 1220
QY 1146 TTACATGAAATTCGATGTCGAAAGAAATTCGCATGCTAAACCTGGAATAGCCACA 1205
Db      |||      |||      |||      |||      |||      |||      |||      |||
1221 -----LysLysSerGluLeuLeuAspAlaLeuAsnLeuGlnIleLysGlu 1234
QY 1206 CTGAAACACCAATACACGAGAAAGAAATAAATCTTTGAGGACATTAAATTTTAAA 1265
Db      |||      |||      |||      |||      |||      |||      |||      |||
1235 LeuLysLysLysAsnGluThrAsnGluAlaSerLeuLeuGluSerIleLysSerValGlu 1254
QY 1266 GAA-----AAGAATGCTGAACCTTCAGATGACCTTAAACTGAAAGAGGAATCATTA 1316
Db      |||      |||      |||      |||      |||      |||      |||      |||
1255 SerGluThrValLysIleLysGluLeuGlnAspGluCysAsnPheLysGlu----- 1271
QY 1317 ACTAAAGGGCATCTCAATATAGTGGGAGCTTAAAGTTCTGATAGCTGACAAACACAATG 1376
Db      |||      |||      |||      |||      |||      |||      |||      |||
1272 ---LysGluValSerGluLeuGluAspLysLeuLysAlaSerGluAspLysAsnSerLys 1290
QY 1377 CTC-----ACTTCTAAATTGAGGAAACAAAGACAAAGAAATACTA 1418
Db      :|:      :|:      :|:      :|:      :|:      :|:      :|:      :|:
1291 TyrLeuGluLeuGlnLysGluSerGluLysIleLysGluGlnLeuAspAlaLysThrThr 1310
QY 1419 GAGGAGAAATGAATCACCATCCTAGACTGGCTTCTGCTGACAGACCATCATCAA 1478
Db      |||      |||      |||      |||      |||      |||      |||      |||
1311 GluLeuLysIleGlnLeuGluLysIleThrAsnLeuSerLysAlaLysGluLysSerGlu 1330
QY 1479 ATTGACATCAAGAAAGAGTCAGAACCTGCTTCCACATTCGAGGAGATGCTTGTTG 1538
Db      |||      |||      |||      |||      |||      |||      |||      |||
1331 SerGluLeuSerArgLeuLysLysThrSerSer----- 1341
QY 1539 CAAGAGAAATGAATGTTGATGTAGTAGTACCGATATATAACATGAGGTG---CTCCA 1595
Db      :|:      :|:      :|:      :|:      :|:      :|:      :|:      :|:
1342 GluGluArgLysAsnAlaGluGlnLeuGluLysLeu-LysAsnGluIleGlnIleLys 1361
QY 1596 TCACACACTTCTGAGGTCAAAGGAAATCCANAAGCCTAAATAATCAATTATGTC 1655
Db      |||      |||      |||      |||      |||      |||      |||      |||
1361 sAsnGlnAlaPheGluLysGluArgLysLeuLeuAsnGluGlySerSer-ThrIleThrG 1381
QY 1656 AGGAGATGCTCTAAGAGAAATACATTTGTTTCAGGAACATGCACAAAGACCAACCGTG 1715
Db      |||      |||      |||      |||      |||      |||      |||      |||
1381 InGluTy-Ser---GluLysIleAsnThrLeuGluAspGluLeuIleArgLeuGlnAsnG 1400
QY 1716 AAACACAGTGTCAAATGAGGAGCTGAACAC----- 1747
Db      |||      |||      |||      |||      |||      |||      |||      |||
1400 LuAsnGluLeuLysAlaLysGluIleAspAsnThrArgSerGluLeuGluLysValSerL 1420
QY 1748 -----ATGNTTCAANCGACACAGATNATGTGACAAACACACTGANCAGC 1793
Db      :|:      :|:      :|:      :|:      :|:      :|:      :|:      :|:
1420 euSerAsnAspGluLeuGluGluLysGlnAsnThrIleLys-----SerLeuGlnA 1438
QY 1794 AGGAGTCTCTA-----GATCAGAAATATTTCAACTAC 1826
Db      |||      |||      |||      |||      |||      |||      |||      |||
1438 spGluIleLeuSerTyrLysAspLysIleThrArgAsnAspGluLysLeuLeuSerIleG 1458
QY 1827 AAGCAAAATATG-----TGCTTCAACAGCAATATTCTTCATGCACAT---- 1870
Db      :|:      :|:      :|:      :|:      :|:      :|:      :|:      :|:
1458 luArgAspAsnLysArgAspLeuGluSerLeuLysGluGlnLeuArgAlaIaGlnGluS 1478
QY 1871 -----AANGAAGCTGCACAAACAAAGCAAGATAACAATTTGATTTTCATTNT 1916
Db      |||      |||      |||      |||      |||      |||      |||      |||
1478 erLysAlaLysValGluGluGlyLeuLysLysLeuGluGluSerSerLysGluLysA 1498
QY 1917 CTTGAGAGGAAATGCNCATCATCTTCTTAAAGAGAAATAAGGAGATATTTNATTACN 1976
Db      :|:      :|:      :|:      :|:      :|:      :|:      :|:      :|:
1498 laGluLeuGluLysSerLysGluMetMetLysLysLeuGluSerThrIleGluSerAsnG 1518
QY 1977 ATAACCATTTAAACACCGGTATATTTCATATATGGAAGAAAAAANAANA 2026
Db      |||      |||      |||      |||      |||      |||      |||      |||
1518 luThrGluLeuLysSerSerMetGluThrIleArgLysSerAspGluLys 1534
```


| | | | | | | |
|----|-------|-----|------|---|--------|-----------------------------|
| 1 | 258 | 7.6 | 1358 | 2 | UC5421 | smooth muscle myosin |
| 15 | 258 | 7.6 | 1358 | 2 | JC5420 | smooth muscle myosin |
| 16 | 258 | 7.6 | 1358 | 2 | UC5420 | smooth muscle myosin |
| 17 | 256.5 | 7.5 | 1790 | 2 | S67593 | transport protein |
| 18 | 255.5 | 7.5 | 2442 | 2 | T09621 | centrosome associated |
| 19 | 255 | 7.5 | 1979 | 1 | S03166 | myosin heavy chain |
| 20 | 255 | 7.5 | 2101 | 2 | A42184 | myosin heavy chain |
| 21 | 254 | 7.5 | 1961 | 1 | A61231 | nuclear mitotic apparatus |
| 22 | 254 | 7.5 | 2245 | 2 | T18278 | myosin heavy chain |
| 23 | 253 | 7.4 | 1939 | 2 | I49175 | myosin heavy chain |
| 24 | 253 | 7.4 | 2253 | 2 | T30336 | nuclear/mitotic apparatus |
| 25 | 252 | 7.4 | 1939 | 1 | A48762 | myosin alpha heavy chain |
| 26 | 250.5 | 7.4 | 1992 | 2 | T46337 | hypothetical protein |
| 27 | 250.5 | 7.4 | 1388 | 2 | S74245 | serine/threonine-s |
| 28 | 249 | 7.3 | 1938 | 1 | S06005 | myosin alpha heavy chain |
| 29 | 248.5 | 7.3 | 853 | 2 | T51505 | myosin alpha heavy chain |
| 30 | 248.5 | 7.3 | 2168 | 2 | T30171 | hypothetical protein |
| 31 | 247.5 | 7.3 | 3488 | 2 | T34418 | ninein - mouse |
| 32 | 247 | 7.3 | 1938 | 2 | I49464 | hypothetical protein |
| 33 | 247 | 7.3 | 1976 | 2 | A59252 | alpha cardiac myosin |
| 34 | 247 | 7.3 | 3259 | 1 | A56539 | myosin heavy chain |
| 35 | 246 | 7.2 | 924 | 2 | S06117 | giantin - human |
| 36 | 245 | 7.2 | 3225 | 2 | I52300 | myosin heavy chain |
| 37 | 245 | 7.2 | 1410 | 1 | A57013 | giantin - human |
| 38 | 244 | 7.2 | 1130 | 2 | T34081 | early endosome antigen |
| 39 | 244 | 7.2 | 1295 | 2 | T24587 | hypothetical protein |
| 40 | 243.5 | 7.2 | 1935 | 1 | A37102 | hypothetical protein |
| 41 | 243 | 7.1 | 1085 | 2 | P96712 | myosin beta heavy chain |
| 42 | 243 | 7.1 | 1999 | 1 | S21801 | hypothetical protein |
| 43 | 242.5 | 7.1 | 1392 | 2 | A43336 | myosin heavy chain |
| 44 | 242 | 7.1 | 1369 | 2 | F84730 | microtubule-vesicle |
| 45 | 242 | 7.1 | 3187 | 2 | T49330 | probable myosin heavy chain |
| 46 | 242 | 7.1 | 3187 | 2 | UC5837 | 364kGolggi complex |

```

ALIGNMENTS

RESULT 1
T14867
Interaptin - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C:Accession: T14867
R:Rivero, F.J.; Kuspa, A.; Brokamp, R.; Matzner, M.; Noegel, A.A.
J. Cell Biol. 142. 735-750, 1998

```

ts.
A:Title: Interleukin, an actin-binding protein of the alpha-actinin superfamily in Dicty.
A:Reference number: Z18248; MUID:98365468; PMID:9700162
A:Accession: T14867
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1738 <RIV>
A:Cross-references: EMBL:AF057019; NID:g3549260; PID:g3549261; PIDN:AAC34582.1
C:Genetics:
A:Gene: abcd

```

Alignment Scores:
Pred. No.:      6.89e-12      Length:      1738
Score:          318.50        Matches:      180
Percent Similarity: 37.86%    Conservative: 146
Best Local Similarity: 20.91% Mismatches:    244
Query Match:    9.36%        Indels:       291
DB:             2            Gaps:         33

US-09-602-362E-15 (1-2030) x T14867 (1-1738)

Qy      69  TTAGAAATTCATGCAGATGCAGAAACTTTCAAAGCAGAGCCCTCCGAGAGCCATCTGCCTTC 128
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      585  LeuGlnGlnGlnAspGlnAspGlnAspIleIys-----GluLysGluPheGlnPhe 600
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy      129  GAGCCT-----GCGATTGAAATGCAG 149
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      601  GluLysGlnGlnLeuLeuSerGlnIleAspSerIleThrThrAsnIleGlnGluIyrGln 620
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

| | | | |
|----|------|---|------|
| QY | 150 | AGTCTGTTCCAAATAAAGCCTTGGAAATTGAAGAATGAACAACATTTGAGACGAGATGAG | 209 |
| Db | 621 | AspLysPheAsnAsnLeuGlnGlnGluPheAsnThrGlnGlnThrLeuAsnGlnGlnGlu | 640 |
| QY | 210 | ATA-----CTCCCATCAGAATCCAAACAA-----AAGGACTATCAAGAAAGTTCTTGG | 257 |
| Db | 641 | ThrHisArgLeuThrGlnGlnLeuTyGlnIleAsnThrAspTyrAsnGlnLysGlnThr | 660 |
| QY | 258 | GAT-----TCTGAGAGTCTCTGTGAGACTGTT | 284 |
| Db | 661 | GlnLeuGlnSerGluIleLysAspAsnGlnThrIleAsnGlnGlnLeuAsnLysGlnLeu | 680 |
| QY | 285 | TCACAGAAGGATGTGTGTTTACCACCAAGGCTACACATCAAAAAAGAA-----ATAGAT | 335 |
| Db | 681 | SerGluLysAspLysGluIleGluLysLeuSerAsnGlnGlnGlnGlnGlnAspGlu | 700 |
| QY | 336 | AAATATAATGGAAAAATTAGNAGAGTCTCTGATAATGATGCTTTTCTGAAGGTCCTCCG | 395 |
| Db | 701 | LysIleAsnAsnLeuLeuGluIleLysGluLysAspCysLeuIleGlu----- | 717 |
| QY | 396 | AGAATGAAGTTTCTATTCCCAATAAAGCCTTAGAA----- | 431 |
| Db | 718 | -----ArgIleAsnGlnGlnLeuLeuGluAsnIleAspLeuAsnSerLysTyr | 733 |
| QY | 432 | -----TTGATGGACATGCAAACTTTCAACAGCAGAGCCTCCGAGAAGCCATCTGCC | 482 |
| Db | 734 | GlnGlnLeuLeuLeuGluPheGluAsnPheLys----- | 744 |
| QY | 483 | TTGAGCCTCCCATTTGAAATGCAAAATCGTTTCCAAATAAAGCCCTTGGAAATTGAAGAAT | 542 |
| Db | 745 | -----LeuAsnSerSerLysGluLysGluAsnGlnLeuAsnGlnLeuGlnSer | 760 |
| QY | 543 | GAACAAACATTGAGACGACATCAGATGTTCCCT----- | 575 |
| Db | 761 | LysGlnAspGluArgPheAsnGlnLeuAsnAspGluLysGluLysGlnLeu | 780 |
| QY | 576 | -----TCAGAAATCAAAACAAAAGAGGTTGAAGAAAATCTTGG | 614 |
| Db | 781 | GlnSerIleGluAspGluPheAsnGlnTyrLysGlnGlnLeuSerSerAsnSerAsn | 800 |
| QY | 615 | GATTCTGAGATCTCCGTGAGCTGTTTCACAGAAGGATGTGTGTACCCCAAGGCTACA | 674 |
| Db | 801 | IleAspGlnGlnLeuGlnSerThrIleIleGlu-----LeuSerGluLeuLys | 816 |
| QY | 675 | CATCAAAAAGAAATGGAT-----AAATAGTGGAAATTAGAAGATTCAACTAGCCCTA | 728 |
| Db | 817 | GluGlnLysGluLeuAsnAspSerLysLeuIleGluLysGlnLysGlnLeuGlnLeu | 836 |
| QY | 729 | TCAAAAATCTTGATACAGTTCATTCTGTGTGAAGACCAAGGAACTTCAAAAGATCAC | 788 |
| Db | 837 | GlnGlnGluPheAspGlnLeuAsn-----GluLysAsnGlnLysAspHis | 851 |
| QY | 789 | TGTGAACACAGCTACAGGAAAAATGGAAACAAATGAAAAAGAAAGTTTGTGTACTGAAAAG | 848 |
| Db | 852 | GlnAspGln-----LeuGluLeuLysGlnLeuLysGlnLeuGlnGln | 867 |
| QY | 849 | AACTGTCAAGCAACAAAGAAATAAAATACAGTTACAGACCAACAAAGTTAAATGG-- | 905 |
| Db | 868 | GluTyrAspGlnLeuAsnGluThrAsnGlnSerIleGluAsnGlnLeuAsnGlnGlnAsn | 887 |
| QY | 906 | -----GAACAAGAGCTCTGCAGTGTGAGATGGAATGAACT | 935 |
| Db | 888 | LeuIleAsnLysGluAsnLeuAsnGluLysGlnGlnLeuLysLeuGlnAsnGln | 907 |
| QY | 936 | TTAAACCAAGACAGACAGACAGAGA----- | 962 |
| Db | 908 | LeuAsnGlnGlnIleGluLysIleGlnPheAspGlnGlnGluPheSerLysGlnAsnSer | 927 |
| QY | 963 | ---AATGCCGATATATTAAATGAAAA-----ATTAGGAGAAATTA | 1001 |
| Db | 928 | IleAsnIleGluLeuValAsnGluLysAsnGluLysLeuIleGlnLeuGlnGlnAspTyr | 947 |
| QY | 1002 | CGAGAGATCGAAGACGACATAGG----- | 1025 |

| | | | |
|----|------|--|------|
| Db | 948 | AspGlnLeuLysGlnGlnAsnArgSerAsnAspGluLysAspGluAsnAspLeuIleGlu | 967 |
| QY | 1025 | ----- | 1025 |
| Db | 968 | LysGluAsnGlnLeuLysSerIleGlnAsnGluLeuAsnGlnLeuIleGluLysAsnGlu | 987 |
| QY | 1026 | -----AAAGAGTTGAAGTGAACAACAA--CTTGAACAGGCTCTCAGAATACAA | 1073 |
| Db | 988 | SerAspHisLysGluGlnLeuLysGlnSerIleGluAsnAspLeuIleGluLys | 1007 |
| QY | 1074 | GATATAGAAATTGAAGAGTGTAGAACTAATTG-----AAT | 1109 |
| Db | 1008 | GluAsnGlnIleGlnGlnLeuGlnSerGlnLeuAsnGluGlnArgGlnGlnGlnSerAsn | 1027 |
| QY | 1110 | CAGTTTCTCACACTCATGAAATGAAATTAATCTCTTACAT----- | 1151 |
| Db | 1028 | GlnLeuSerGluLysAspGlnGlnLeuAsnGlnLeuIleGluLysAsnGlnPheAspGln | 1047 |
| QY | 1152 | -----GAAATTCATGTTTGAAGAGGAATTGCC | 1181 |
| Db | 1048 | LysGluGlnGlnLeuLysGlnSerIleGluAsnAspLeuPheGluLysGluAsnGln | 1067 |
| QY | 1182 | ATCTAAACTCGAAATAGCCACCTGAAACACCAATACACAGAAAGAGAAATAAATAC | 1241 |
| Db | 1068 | IleGlnGlnLeuGln--SerGlnLeuAsnGluGlnArgGlnGlnSerAsnGlnLeu | 1086 |
| QY | 1242 | TTTGAG-----GACATTAGATTTTAAAGAAAGATCTCAACTTCAGATGACC | 1292 |
| Db | 1087 | SerGluLysAspGlnGlnLeuAsnGlnLeuIleGluLysAsnGlnSerAspGlnLysGlu | 1106 |
| QY | 1293 | CTAAACTGAACAGAAATCA-----TTAACTAAAGGCGATCTCAATATAGT | 1340 |
| Db | 1107 | GlnGlnLeuLysGlnGlnSerIleGluAsnAspLeuIleGluLysGluAsnGlnIleGln | 1126 |
| QY | 1341 | GGCAGCTTAAAGTTCTATAGTGTAGACACAAATGCTCACTTCTAAATG----- | 1391 |
| Db | 1127 | --GlnLeuGlnLeuGlnLeuAsnGluGlnArgGlnLeuGlnSerGluValSerIleAsp | 1145 |
| QY | 1391 | ----- | 1391 |
| Db | 1146 | AsnAspLysIleLeuGluLeuGluLysGlnLeuLysGlnCysGlnSerAspLeuLys | 1165 |
| QY | 1392 | -----AAGAAAAACAAGCAAGAATACTAGAGGAGAAATTGAA----- | 1433 |
| Db | 1166 | LeuAsnAspGluLysGlnGlnAspLysGlnLeuGlnAspLysGlnIleGluPheAsp | 1185 |
| QY | 1433 | ----- | 1433 |
| Db | 1186 | GlnLeuGlnLeuThrPheAsnGlnPheLysAsnAspLysAspSerGlnPheIleGlnLeu | 1205 |
| QY | 1434 | -----TCACACCATCTAGACTGGCTTCTGTGTACAGACCATG-----ATCAAT | 1481 |
| Db | 1206 | GlnAspAspGlnLysGlnGlnLeuGlnSerIleGlnGlnAsp-LeuAsnGlnLeuLysGln | 1225 |
| QY | 1482 | GTGCATCTAGAAAAAGTCAAGAACCTG-----CTTTCCACATTGCA | 1523 |
| Db | 1225 | nGluAsnGlnGluLysGlnLeuSerGlnLeuSerGluLysAspGluLysLeuGlnSerIleGln | 1245 |
| QY | 1524 | GGAGATCTTTTTCAGAAAAAATGAATGTTGATGTAGTAGTACCGATATATAACAA | 1583 |
| Db | 1245 | n-----PheGluAs | 1248 |
| QY | 1584 | TGAGTGCTCCATCAACCATCTTTCTGAAGCTCAAGGAATCCANAAGCCTCAAAATTA | 1643 |
| Db | 1248 | nGlnGluLysGlnLysGlnLeuSerGluLysAspGluLysLeuGlnSerIleGlnGlnAs | 1268 |
| QY | 1644 | TCTCAATATTGCAGAGATGCT---CTAAGAGAAAAATACATTGTTTCAGAA- | 1693 |
| Db | 1268 | nLeuAsnGlnLeuAsnAspGluAsnGln-GluLysValLysGlnPheSerGlnLysAspGln | 1288 |
| QY | 1694 | -----CATGCACAAAGAGACCAACGCTGAAACAC | 1721 |

Db 1288 lyllysleuGlnSerIleGlnGlnAspLeuAsnGlnLeuLysGlnGluAsnGlnGluLysG 1308
 Qy 1722 AGTGTCAATGAAGAGCTGAACACATGNTCAAACGAACAGATNATGTGAACAAAC 1781
 Db 1308 lyllysGlnLeuSerGluLysAspGluLysLeuGlnSerIleGlnGlnAspLeuAsnGlnL 1328
 Qy 1782 ACAGTGNACAGCAGGAGCTCTAGATCAGAAATATTTCAACTACAAAGCAAAATATGT 1841
 Db 1328 euAsnAspAspGlnIleLysLysAsnGluLysLeuLysGluLysGluGlnLeuLeuL 1348
 Qy 1842 CGGTTCAACAGCAATAGTTTCATGCATAAAGAAAGCTCACAAACAAAGCAAGATAACA 1901
 Db 1348 ylsleuGlnGlnAspPheAsnAspGlnSerGlnGlnLeuLysGlnLeuGluGlu---- 1366
 Qy 1902 ATTGATNTTCATNTTCTTGAGAGAAATGCNCATCATCTTCTAAAGAGAAATATGAG 1960
 Db 1367 -----lysLeuSerGluLysGluAsnGlnLeuGlnGlnLeuLysGlnGluAsnGlu 1383
 RESULT 2
 S38173
 myosin-like protein MLPI - yeast (Saccharomyces cerevisiae)
 N;Alternate names: protein YKR095w; protein YKR415
 C;Species: Saccharomyces cerevisiae
 C;Date: 03-May-1994 #sequence revision 03-May-1994 #text_change 21-Jul-2000
 C;Accession: S38173; S40647; S31207
 R;Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo, J.
 submitted to the Protein Sequence Database, March 1994
 A;Reference number: S38158
 A;Accession: S38173
 A;Molecule type: DNA
 A;Residues: 1-1875 <BAL>
 A;Cross-references: EMBL:Z28320; NID:G486586; PID:G486587; MIPS:YKR095w
 R;Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; J.
 Yeast 9, 1349-1354, 1993
 A;Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chromo
 A;Reference number: S40644; MUID:94205265; PMID:8154186
 A;Accession: S40647
 A;Molecule type: DNA
 A;Residues: 1-1875 <BOU>
 A;Cross-references: EMBL:X73541; NID:G450550; PIN:CAA51948.1; PID:G450554
 A;Experimental source: strain S288C
 R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
 Mol. Gen. Genet. 237, 359-369, 1993
 A;Title: A new yeast gene with a myosin-like heptad repeat structure.
 A;Reference number: S31207; MUID:93247549; PMID:8483450
 A;Accession: S31207
 A;Molecule type: DNA
 A;Residues: 1-300, A', 302-1875 <KOE>
 A;Cross-references: EMBL:L01992; NID:G171958; PIDN:AAA34783.1; PID:G171959
 C;Genetics:
 A;Gene: SGD:MLP1
 A;Cross-references: SGD:S0001803; MIPS:YKR095w
 A;Map position: 11R
 Alignment Scores:
 Pred. No.: 1.33e-09 Length: 1875
 Score: 280.50 Matches: 181
 Percent Similarity: 40.05% Conservative: 145
 Best Local Similarity: 22.24% Mismatches: 277
 Query Match: 8.24% Indels: 211
 DS: 2 Gaps: 39
 US-09-602-362E-15 (1-2030) x S38173 (1-1875)
 Qy 33 AACTCGGAATGAAAGTTCTTATT--CCAACTAAAGCCTTAGAATTGATGGACATGCAA 89
 Db 737 SerCysLysSerLysLeuSerIleValGluThrGluLeuLeuAsnLeuLysGluGluGln 756
 Qy 90 ACTTTCAAAGCAGAG-----CCTCCCGAG 113
 Db 757 LysLeuArgValHisLeuGluLysAsnLeuLysGlnGluLeuAsnLysLeuSerProGlu 776

Qy 114 AAGCCATCTGCCTTCGAGCCTGCCATTGAAATGCAAAAGTCTGTTCAAATAAAGCCTTG 173
 Db 777 LysAspSerLeuArgIleMetValThrGlnLeuGln-----ThrLeu 790
 Qy 174 GAATTGAAGAATGAACAAACATTG-----AGAGCAGAT 206
 Db 791 GlnLysGluArgGluAspLeuGluGluThrArgLysSerCysGlnLysLysIleAsp 810
 Qy 207 GAGATA-----CTCCATCAGAAATCCAAACAAAGAGATAT----- 242
 Db 811 GluLeuGluAspAlaLeuSerGluLeuLysGluThrSerGlnLysAspHisIle 830
 Qy 243 -----GAAGAAAGTTCT-----TGG----- 257
 Db 831 LysGlnLeuGluGluAspAsnSerAsnIleGluIleThrGlnAsnLysIleGluAla 850
 Qy 258 -----GATTCGAGAGTCTCTGTGAGACTGTTTCACAGAGAGTGTGTGTTATCCC 308
 Db 851 LeuLysLysAspTyrGluSerValIleThrSerValAspSerLysGlnThrAspIleGlu 870
 Qy 309 AAGGCTACACAT-----CAAAAAGAAATA-----GATAAATAAAT----- 344
 Db 871 LysLeuGlnTyrLysValLysSerLeuGluLysGluIleGluAspLysIleArgLeu 890
 Qy 345 -----GGAAATTAGAAGACTCTCTGATATGATGCTTTTCTCAAGGCTCCCTGCG 395
 Db 891 HisThrTyrAsnValMetAspGluThrIleAsnAspAspSerLeuArgLysGluLeuGlu 910
 Qy 396 AGAATGAAGTTTCTATTCCA-----ACTAAAGCCTTAGAATTGATGGACATG--- 443
 Db 911 LysSerLysIleAsnLeuThrAspAlaTyrSerGlnIleLysGluTyrLysAspLeuTyr 930
 Qy 444 -----CAAACTTTCAAGCAGAGCTCCCGAAGAGCCATCTGCTTCGAGCCT 491
 Db 931 GluThrThrSerGlnSerLeuGlnGlnThrAsnSerLysLysLeuAspGluSerPheLysAsp 950
 Qy 492 GCCATTGAAATGCAAAAGTCTGTCCAAATAAAGCCTTGAAATTGAAGAATGAACAACA 551
 Db 951 PheThrAsnGlnIleLysAsnLeuThrAspGluLysThrSerLeuGluAspLysIleSer 970
 Qy 552 TTGAGAGCAGATCAGATGTTTCCCTTCAGAAATCAAAA-----CAAAAGAGAGTTGAA 602
 Db 971 LeuLeuLysGluGlnMetPheAsnLeuAsnAsnGluLeuAspLeuGlnLysLysGlyMet 990
 Qy 603 GAAATTTCTGGGATTCAGAGCTCTGAGAGTCTCCGTGAGACTGTTTCACAG-----AAGGAT 653
 Db 991 GluLysGluLysAlaAspPheLysLysArgIleSerIleLeuGlnAsnAsnLysGlu 1010
 Qy 654 GTGTGTGTACCCCAAGGCTACACATCAAAAGAAATGATAAATAAGTGGAAATTAGAA 713
 Db 1011 ValGluAlaValLysSerGluTyrGluSerLysLeuSerLysIleGlnAsnAspLeuAsp 1030
 Qy 714 GATTCAACT----- 722
 Db 1031 GlnGlnThrIleTyrAlaAsnThrAlaGlnAsnAsnTyrGluGlnGluLeuGlnLysHis 1050
 Qy 723 -----AGCCTATCAAAATCTGGATACATGATTCATCTCTGTGAA----- 761
 Db 1051 AlaAspValSerLysThrIleSerGluLeuArgGluGlnLeuHisThrTyrLysGlyGln 1070
 Qy 762 ---AGAGCAAGGGAATCTCAAAAGATCACTGTGAACAACGTCACAGGAAAAATGGAACAA 818
 Db 1071 ValLysThrLeuAsnLeuSerArgAspGlnLeuGlu-----AsnAlaLeuLysGlu 1087
 Qy 819 ATGAAAAAGAGTTTGTGTACTGAAAAAGAACTCTCAGAGCAAGCAAAATA---AAA 875
 Db 1088 AsnGluLysSerTrpSerGlnLysGluSerLeuLeuGluGlnLeuAspLeuSerAsn 1107
 Qy 876 TCACAGTTAGAGAAC-----CAAAAGTTAAATGGGAACAA----- 911
 Db 1108 SerArgIleGluAspLeuSerSerGlnAsnLysLeuLysLeuTyrAspGlnIleGlnIleTyr 1127
 Qy 912 -----GAGCTCTGCAGTGTGAGA 929

QY 357 ---GAGTCTCCTGAT-----AAT 371
 Db 1315 AlaGluAsnAlaAspLeuAlaThrGluLeuArgSerValAsnSerArgGlnGluAsn 1334
 QY 372 GATGGTTTCTGAAGGCTCCCTGCAGATGAAGTTCTTATTCCAACTAAAGCCTTAGAA 431
 Db 1335 AspArgArgGlys-----GlnAlaGluSerGlnIleAlaGluLeuGlnVallys 1351
 QY 432 TTGATGGACATCAAACTTTCAAGCAGAGCCTCCCGAGAGCCATCTGCCTTCGAG--- 488
 Db 1352 LeuAlaGluIleGluArgAlaArgSerGluLeuGlnGluLysCysThrLysLeuGlnGln 1371
 QY 489 -----CCTGCCATTGAATGCACAAAGCTCTGTTCCAAAT 521
 Db 1372 GluAlaGluAsnIleThrAsnGlnLeuGluGluAlaGluLeuLysAlaSerAlaAlaVal 1391
 QY 522 AAAGCCTTGGAAATTGAAGAAATGAACAAACATTGAGAGCAGATCAGATGTTCCCTTCAGAA 581
 Db 1392 LysSerAlaSerAsnMetGluSerGlnLeuThrGluAlaGlnGlnLeuLeuGluGluGlu 1411
 QY 582 TCAAAACAAAAG-----AAGTTGAAGAAATCTTGGATCTCGAGAT 626
 Db 1412 ThrArgGlnLysLeuGlyLeuSerSerLysLeuArgGlnIleGluSerGluLysGluAla 1431
 QY 627 CTCCGTGAGACTGTTTTCACAGAGAGGATGTGTGTACCCAGGCTACACATCAAAAA--- 683
 Db 1432 LeuGlnGluGlnLeuGluGluAspAspGluAla-----LysArgAsnTyrGluArgLys 1449
 QY 684 -----GAATGATAAATAAGTGAATAATTAAGATTAAGATTAAGTAC 725
 Db 1450 LeuAlaGluValThrGlnMetGlnGlnIleLysLysAlaGluGluAspAlaAsp 1469
 QY 726 CTATCAAAATCTTGGATACAGTTTCATCTTGTGAAGAGAGGAACTTCAAAAGAT 785
 Db 1470 LeuAlaLysGluLeuGlu-----GluGlyLysLysArgLeuAsnLysAsp 1484
 QY 786 CACTGTGAACACGTACAGGAAAATGAACAAATGAATAAGTGTGTACTGAAA 845
 Db 1485 -----IleGluAlaLeuGluArgGlnValLysGluLeuIleAlaGlnAsn 1499
 QY 846 AAGAACTGTCAGAGCA---AAGAAATAAATCAGATGTAGAGCAACCAAAAGTTAAA 902
 Db 1500 AspArgLeuAspLysLysLysLysLysLysGlnSerGluLeuGluAspAlaThrIleGlu 1519
 QY 903 TGGGAACAGAGCTCTGCAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAGA 962
 Db 1520 LeuGluAlaGln-----ArgThrLysValLeuGluLeuGluLysLysGlnLys 1535
 QY 963 AATGCCGAT-----ATATTAAATGAATAAATTAGGAGAAATTA 1001
 Db 1536 AsnPheAspLysIleLeuAlaGluGluLysAlaIleSerGluGlnIleAlaGlnGluArg 1555
 QY 1002 GGAAGAAATCGAA-----GAGCAGCATAGGAAAGAGTTAGAGTCAAAACAA 1049
 Db 1556 AspThrAlaGluArgGluAlaArgGluLysGluThrLysValLeuSerValSerArgGlu 1575
 QY 1050 CTTGAACAGGCTCTC---AGAATCAAGATATAGAA-----TTCAAGAGTGTAGAAAGT 1100
 Db 1576 LeuAspGluAlaPheAspLysIleGluAspLeuGluAsnLysArgLysThrLeuGlnAsn 1595
 QY 1101 AATTTGAATCAGGTTTCTCACACTCATGAAATGAATTAATCTTACATGAAATTCG 1160
 Db 1596 GluLeuAspAspLeuAlaAsnThrGlnGlyThrAlaAspLysAsnValIleGlu----- 1613
 QY 1161 ATGTTGAAAGGAAATTTGCCATGCTAAATCTGGAATAGGCACACTGAAACACCAATAC 1220
 Db 1614 ---LeuGluLysAlaLysArgAlaLeuGluSerGlnLeuAlaGluLeuLysAlaGlnAsn 1632
 QY 1221 CAGGAAAGGAAAT-----AAATACTTTGAGGACATTAGATTTTAAAGAAAGAAAT 1274
 Db 1633 GluGluLeuGluAspAspLeuGlnLeuThrGluAspAlaLysLeuArgLeuGluValAsn 1652

QY 1275 -----GCTGAACCTTCAGATGACCTTAAACTGAAAGAGGAAATCATTAAC 1319
 Db 1653 MetGlnAlaLeuArgSerGlnPheGluArgAspLeuLeuAlaLysGluGluGlyAlaGlu 1672
 QY 1320 AAAAGGCATCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAG----- 1367
 Db 1673 GluLysArgArgGlyLeuValLysGlnLeuArgAspLeuThrGluLeuAspGluGlu 1692
 QY 1368 -----AACCAATGCTCCTTAAATTTGAAGAAACAAAGCAAGAAATA 1415
 Db 1693 ArgLysGlnArgThrAlaAlaValAlaSerLysLysLysLeuGluGlyAspLeuLysGlu 1712
 QY 1416 CTAGAGCAGCAAAATGGAATCACACCATCTAGACTGCTTCTGCTGTACAGACCATGAT 1475
 Db 1713 IleGluThrThrMetGluMetHisAsnLysValLysGluAspAlaLeu-LysHisAlaLys 1732
 QY 1476 CAAATGTCATCAGAAAAAAGTCAAGAACCTGCTTCCACATTCAGAGAGATGCTTGT 1535
 Db 1732 sLysLeu-----GlnAlaGlnValLysAspAlaLeuArgAspAlaGluGluAlaLysAl 1750
 QY 1536 TTGCAAGAAATAATGAATGTTGATG-----TGAGTAGTACCGATATATA 1580
 Db 1750 AlaLysGluGluLeuGlnAlaLeuSerLysGluAlaAspGlyLysValLysAlaLeuGlu 1770
 QY 1581 CAATGAGCTGCTCCAT-----CAACCATCTTCTGAAGCTCAAGGAAATCCANAAGCCT 1634
 Db 1770 uAlaGluValLeuGlnLeuThrGluAspLeuAlaSerSerGluArgAlaArgAlaAl 1790
 QY 1535 AAAAATTATCTCAATATGTCAGAGAGATGCTTAAAGAAATAATACATTTGTTTCAGAAC 1694
 Db 1790 aGluThrGluArgAspLeuAlaGluGluIleAlaAsnAsnAlaAsnLysGlySerLe 1810
 QY 1695 ATGCACAAAGAGACCAACGTCGAAACAGTCGTCAAATGAAGGAGCTGACACATGNTC 1754
 Db 1810 uMetIleAsp-GluLysArgArgLeuGluAlaArgIleAlaThrLeuGluGluLeuG 1830
 QY 1755 AAANCGAACAA-----GATNATGTGAACAACAACACACTGANCAGC 1793
 Db 1830 luGluGluGlnSerAsnSerGluValLeuLeuAspArgSerArgLysAlaGlnLeuGlnI 1850
 QY 1794 AGGAGTCTCTAGATCAGAAATTTTCACTACAAAGCAAAAATATGTGGCTTCAACAGC 1853
 Db 1850 leGluGlnLeuThrGluLeu----- 1857
 QY 1854 AATTAGTTCATGCACATAAAGCTGACAAACAAAGCAAGATAACAATTTGATNTTCAT 1913
 Db 1858 -----AlaAsnGluLysSerAsnSerGlnLys----- 1866
 QY 1914 TMTCTTGAGGAGAAATGCNCATCATCTTCTTAAAGAGAAATAATGAGGAGATA 1966
 Db 1867 -----AsnGluAsnGlyArgAlaLeuLeuGluArgGlnAsnLysGluLeu 1881
 RESULT 4
 A26655
 myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)
 N/Contains: myosin ATPase [EC 3.6.4.1]
 C/Species: Dictyostelium discoideum
 C/Date: 05-Oct-1988 #sequence revision 05-Oct-1988 #text_change 19-Apr-2002
 C/Accession: A26655; A24728; S00250
 R/Warrick, H.M.; De Lozanne, A.; Levinwand, L.A.; Spudich, J.A.
 Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986
 A/Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium disco
 A/Reference number: A26655; MUID:87092266; PMID:3540939
 A/Accession: A26655
 A/Molecule type: DNA
 A/Residues: 1-2116 <WAR>
 A/Cross-references: GB:M14628; GB:M11938; NID:g167834; PID:AAA33227.1; PID:g167835
 R/DeLozanne, A.; Lewis, M.; Spudich, J.A.; Levinwand, L.A.
 Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985
 A/Reference number: A24728; MUID:86016788; PMID:3901008
 A/Accession: A24728
 A/Molecule type: mRNA
 A/Residues: 2035-2116

R:Wagle, G.; Noegel, A.; Scheel, J.; Gerisch, G.
 PDB: 1ett, 227, 71-75, 1988
 A:Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostelium
 A:Reference number: S00250; PMID:188112226; PMID:2828113
 A:Accession: S00250
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1734-1993 <WAG>
 C:Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; phosphatase
 F:1-818/Domain: globular head <HD>
 F:89-747/Domain: myosin motor domain homology <MMOT>
 F:179-186/Region: nucleotide-binding motif A (P-loop)
 F:819-2116/Domain: alpha-helical rod <ROD>

Alignment Scores:
 Pred. No.: 1.43e-09 Length: 2116
 Score: 280.00 Matches: 158
 Percent Similarity: 36.61% Conservative: 110
 Best Local Similarity: 21.58% Mismatches: 256
 Query Match: 8.23% Indels: 208
 DB: 2 Gaps: 27

US-09-602-362E-15 (1-2030) x A26655 (1-2116)

QY 24 CTGAAGGCTAACTGCGAATGAAGTTCTATTCCAACCTAAAGCCCTTGAATG---ATG 80
 |||||
 Db 879 LeuLysAlaMetTyrAspSerLysAspAlaLeuGluAlaGlnLysArgGluLeuGluile 898
 |||||
 QY 81 GACATCAAACTTCAAGCAGAGCTCCCGAGAGCCATCTGCCTTCGAGCCTCCCAT 140
 |||||
 Db 899 ArgValGluAspMetGluSerGluLeuAspGluLysLeuAlaLeuGlnLeuGln 918
 |||||
 QY 141 GAAATCAAAAGTCTCTCCAAATAAAGCCTTG-----GAATTGAAGAAATGAA 188
 |||||
 Db 919 AsnGlnLysArgSerValGluLysValArgAspLeuGluGluGluLeuGluGlu 938
 |||||
 QY 189 CAACATCTGAGCAGATGATCTCCCATCAGATCCAAACAAAGACTATCAAGAA 248
 |||||
 Db 939 GlnLysLeuArgAsn-----ThrLeuGluLysLeuLysLysLysTyrGluGlu 954
 |||||
 QY 249 AGTCTCTGGAT-----TCTGAGAGTCTCTGTCGAGACTGTTTCACAGAGGAT 296
 |||||
 Db 955 GluLeuGluGluMetLysArgValAsnAspGlyGlnSerAspThrIleSerArgLeuGlu 974
 |||||
 QY 297 GTGTGTTTACCCAGCTCACATCAAAAGAAATAGATAAATAATCGGAAATTA--- 353
 |||||
 Db 975 Lysile-----LysAspGluLeuGlnLysGluValGluGluLeuThrGluSerPheSer 992
 |||||
 QY 354 GAAGAGTCTCTGATATGATGTTTCTGAGGCTCCCTGCAGAGTAAAGTCTTCTATT 413
 |||||
 Db 993 GluGluSerLysLysLysGlyValLeuGluLysThrArgValArgLeuGlnSerGluLeu 1012
 |||||
 QY 414 CCAACTAAAGCCCTTAGAATTGATGACATGCAAACTTTCAAAGCAGAG----- 461
 |||||
 Db 1013 AspAspLeuThrValArgLeuAspSerGluThrLysAspLysSerGluLeuLeuArgGln 1032
 |||||
 QY 462 -----CCTCCGAGAGCCATCTGCCTTCGAGCTGCC 494
 |||||
 Db 1033 LysLysLysLeuGluGluLeuLysGlnValGlnGluAlaLeuAlaGluThrAla 1052
 |||||
 QY 495 ATTGAATGCAAAAGTCTGTTCCAAATAAAGCCTTG----- 530
 |||||
 Db 1053 AlaLysLeuAlaGlnGluAlaAlaAsnLysLysLeuGlnGlyGluTyrThrGluLeuAsn 1072
 |||||
 QY 530 ----- 530
 |||||
 Db 1073 GluLysPheAsnSerGluValThrAlaArgSerAsnValGluLysSerLysThrLeu 1092
 |||||
 QY 531 -----GAATTGAAGAAATGAACAAACA 551
 |||||
 Db 1093 GluSerGlnLeuValAlaValAsnAsnGluLeuAspGluGluLysLysAsnArgAspAla 1112
 |||||

QY 552 TTGAGA-----GCAGATCAGATGTTCT 572
 |||||
 Db 1113 LeuGluLysLysLysLysAlaLeuAspAlaMetLeuGluGluMetLysAspGlnLeuGlu 1132
 |||||
 QY 573 CTTTCAGATCAAAACAAAG-----AAGTTCAAGAAATCTTTGGAT 617
 |||||
 Db 1133 SerThrGlyGluLysLysSerLeuTyrAspLeuLysValLysGlnGluSer---ASP 1151
 |||||
 QY 618 TCTGAGAGTCTCCGTGAGAGCTGTTTCACAGAGAGTGTGTGTACCC----- 665
 |||||
 Db 1152 MetGluAlaLeuArgAsnGlnIleSerGluGlnSerThrIleAlaLysLeuGluLys 1171
 |||||
 QY 666 ---AAGCTACATCAAAAGAAATGATATAAATAGTGGAAATAGAAATCAACT 722
 |||||
 Db 1172 IleLysSerThrLeuGluGluValAlaArgLeuGlnGlyGluLeuGlu----- 1188
 |||||
 QY 723 AGCTATCAAAATCTTGGATACAGTTTCATTCTGTGAAAGACCAAGGAACTTCAAAA 782
 |||||
 Db 1189 -----AlaGluGlnLeuAlaLys 1194
 |||||
 QY 783 GATCAGTGTGAACACAGTACAGGAAATCGAA---CAATGAAAGAAAGTTTCTGTA 839
 |||||
 Db 1195 SerAsnValGluLysGlnLysLysValGluLeuAspLeuGluAspLys----- 1211
 |||||
 QY 840 CTGAAAAGAACTGTCTGAGAGCAAAAGAAATAAATCAAGTTAGAGAACCAAAAAGTT 899
 |||||
 Db 1212 ---SerAlaGlnLeuAlaGluThrAlaAlaLysGlnAlaLeuAspLysLysLys 1230
 |||||
 QY 900 AAATGGGAACAAGAGCTCTGAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGA 959
 |||||
 Db 1231 LysLeuGluGlnLeuSerGluValGlnThrGlnLeuSerGluAlaAsnAsnLysAsn 1250
 |||||
 QY 960 AGAATCCCATATATTAATGAAATAATAGGAGAAATAGGAGAAATCGAAGAGCAG 1019
 |||||
 Db 1251 ValAsnSerAspSerThrAsnLysHisLeuGluThrSerPheAsnAsnLeu----- 1267
 |||||
 QY 1020 CATAGGAAAGAGTTAGAAAGTGAACCAACACTTGAACAGCTCTCAGAAATACAGATA 1079
 |||||
 Db 1268 ---LysLeuGluLeuGluAlaGluGlnLysAlaLysGlnAlaLeuLysLysArgLeu 1286
 |||||
 QY 1080 GAATTGAAGAGTGTAGAAATGATTTGATTCAGCTTCTCAGACTCATGAAATGAAAT 1139
 |||||
 Db 1287 GlyLeu-----GluSerGluLeuLysHisValAsnGluGlnLeuGluGluLys 1303
 |||||
 QY 1140 TATCTCTTACATGAAATTCATGTTGAAAGAGAAATTCGCCATGCTAAACTGGAATA 1199
 |||||
 Db 1304 LysGlnLysGluSerAsn-----GluLysArgLysValAspLeuGluLysGluVal 1320
 |||||
 QY 1200 GCCACACTGAAACACCAATACAGGAA----- 1226
 |||||
 Db 1321 SerGluLeuLysAspGlnIleGluGluGluValAlaSerLysLysAlaValThrGluAla 1340
 |||||
 QY 1227 -----AAGGAAAT-----AAATCTTTCAGGACATTAAGATT 1259
 |||||
 Db 1341 LysAsnLysLysGluSerGluLeuAspGluLysArgGlnTyrAlaAspValSer 1360
 |||||
 QY 1260 TTAAGAAAGAAATGCTGAAGTTCAGATGACCTTAACTGAAAGAGGATCATTAAT 1319
 |||||
 Db 1361 SerArgAspLysSerValGluGlnLeuLysThrLeuGlnAlaLysAsnGluLeuArg 1380
 |||||
 QY 1320 AAAAGGCATCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGAACACATGCTC 1379
 |||||
 Db 1381 AsnThrAlaGluGluAlaGluGlyGlnLeu----- 1390
 |||||
 QY 1380 ACTTCTAAATTAAGGAGGAAACACAGCAAAATACTAGAGGACAGAAATGAATCACAC 1439
 |||||
 Db 1391 -----AspArgAlaGluArgSerLysLysLysAlaGluPheAsp----- 1403
 |||||
 QY 1440 CATCTAGAGCTGGCTTCTGCTGTACAGACCATCATCAATTTGACATCAAGAAAAAGT 1499
 |||||
 Db 1404 -----LeuGluGluAlaValLysAsnLeuGluGluThrAlaLysLysValLys 1420
 |||||


```

QY 1500 CAAGAACCCTGCTTCCACATTCGAGGAGATGCTTGTTCGAAGAAAATGAATGTTGAT 1559
Db 1421 AlaGluLysAla-----MetLysLysAlaGluThrAsp 1431
QY 1560 GTGAGTAGTACCGATATATAACAATGAGGTGCTCCATCAACCACTTCTCAAGCTCAAG 1619
Db 1432 TyrArgSerThr-----LysSer 1437
QY 1620 GAATCCANAAGCCTAAAATTAATCTCAATTATGAGGAGATGCTCTAAGAGAAAATAC 1679
Db 1438 GluLeuAspAlaLysAsn-----ValSerSerGluGlnTyr 1450
QY 1680 ATTGGTTTCAGGACATCACAAAGACCAACAGTGAACACAGTGTCAAAATCAAGGAAG 1739
Db 1451 ValGlnIleLysArgLeuAsnGluLeuSerGluLeuArgSerValLeuGluGluAla 1470
QY 1740 CTGAACACATGTTTCAAAACGACACAGATNATGTGAACAAACACACTGANCAGCGAGT 1799
Db 1471 AspGluArgCysAsnSerAlaIleLysAlaLysLysThrAlaGluSerAlaLeuGluSer 1490
QY 1800 CTCTAGATCAGAAATTAATTTCACTACAAAGCAAAATATGTGCTTCAACAGCAATTAG 1859
Db 1491 Leu-----LysAspGluIleAspAlaAlaAsnAsn----- 1500
QY 1860 TTCATGCACATGAAGAACTGACACAAACAAAG 1895
Db 1501 -----AlaLysAlaLysAlaGluArgLysSerLys 1510

RESULT 5
A36014
Mycosin heavy chain, nonmuscle - fruit fly (Drosophila melanogaster)
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Drosophila melanogaster
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C:Accession: A36014; B36014
R:Ketchum, A.S.; Stewart, C.T.; Stewart, M.; Kiehart, D.P.
Proc. Natl. Acad. Sci. U.S.A. 87, 6316-6320, 1990
A:Title: Complete sequence of the Drosophila nonmuscle myosin heavy-chain transcript: cDNA
A:Reference number: A36014; MUID:90349606; PMID:2117279
A:Accession: A36014
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-2017 <KET>
A:Cross-references: GB:M35012
C:Genetics:
A:Gene: FlyBase:zfp
A:Cross-references: FlyBase:FBgn0005634
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
F:1-2017/Product: myosin heavy chain, form I #status predicted <MA1>
F:46-2017/Product: myosin heavy chain, form II #status predicted <MA2>
F:135-815/Domain: myosin motor domain homology <MMOT>
F:228-232/Region: nucleotide-binding motif A (p-loop)
F:598-631/Region: actin binding #status predicted
F:705-727/Region: actin binding #status predicted
F:888-2017/Domain: coiled coil #status predicted <COI>
F:888-1328/Region: S2
F:1329-2017/Region: light meromyosin
F:176/Modified site: N6,N6-trimethyllysine (lys) #status predicted
F:231/Binding site: ATP (lys) #status predicted
F:745,755/Active site: Cys #status predicted

Alignment Scores:
Pred. No.: 1,64e-09 Length: 2017
Score: 279.00 Matches: 157
Percent Similarity: 41.30% Conservative: 147
Best local Similarity: 21.33% Mismatches: 262
Query Match: 8.20% Indels: 170
DB: 1 Gaps: 29

US-09-602-362E-15 (1-2030) x A36014 (1-2017)
QY 81 GACATGCAAACTTTCAAAGCAGAGCTCCGAGAGCAATCTGCCTTCGAGCCTGCCATT 140

```

```

Db 1170 AspLeuGluAlaGluLysAla-----AlaArgAlaLysAlaGlu 1182
QY 141 CAATGCAAAAGTCTGTTCCTCAAAATAAGCCCTTGAATTTGAAGAATGAA----- 188
Db 1183 LysValArgAspLeuSerGluGluLeuGluAlaLeuLysAsnGluLeuLysAspSer 1202
QY 189 ---CAACATTGAGAGCAGATGAGATCTCCATCAGATCCAAACAA----- 233
Db 1203 LeuAspThrThrAlaAlaGlnGlnGluLeuArgSerLysArgGluGlnGluLeuAlaThr 1222
QY 234 ---AAGGACTATGAAGAAAGTTCTTGGGATCTCGAGAGTCTCTGAGACTGTTTCA 287
Db 1223 LeuLysLysSerLeuGluGluGluThrValAsnHisGluGlyValLeuAlaAspMetArg 1242
QY 288 CAGAAGGATGTGTTTACCCAGGCTACACATCAAAAGAAATAGATAAAATAATGGA 347
Db 1243 HisLys-----HisSerGlnGluLeuAsnSerIleAsnAsp 1254
QY 348 AAATTAGAA----- 356
Db 1255 GlnLeuGluAsnLeuArgLysAlaLysThrValLeuGluLysAlaLysGlyThrLeuGlu 1274
QY 357 ---GAGTCTCTCGAT-----AAT 371
Db 1275 AlaGluAsnAlaAspLeuAlaThrGluLeuArgSerValAsnSerSerArgGlnGluAsn 1294
QY 372 GATGCTTTTTCGAGGCTCCTCGAGAAATGAAGTTTCTATTTCCTCACTAAAGCCTTAGAA 431
Db 1295 AspArgArgLys-----GlnAlaGluSerGlnIleAlaGluLeuGlnValLys 1311
QY 432 TTGATGCACATGCAAACTTTTCAAGCAGAGCCTCCGAGAAAGCCATCTGCTCTTCGAG-- 488
Db 1312 LeuAlaGluIleGluArgAlaArgSerGluLeuGlnGluLysCysThrLysLeuGlnGln 1331
QY 489 -----CCTGCCATTGAAATGCAAAAGTCTGTTCCTCAAT 521
Db 1332 GluAlaGluAsnIleThrAsnGlnLeuGluGluAlaGluLeuLysAlaSerAlaAlaVal 1351
QY 522 AAAGCCTTGAATTTGAAGATGAACAAACATTCGAGAGCAGATCAGATGTTCCCTTCAGAA 581
Db 1352 LysSerAlaSerAsnMetGluSerGlnLeuThrGluAlaGlnGlnLeuLeuGluGlu 1371
QY 582 TCAAAACAAAG-----AAGTTTGAAGAAAATTTCTGGGATCTCGAGAGT 626
Db 1372 ThrArgGlnLysLeuGlyLeuSerSerLysLeuArgGlnIleGluSerGluLysGluAla 1391
QY 627 CTCGCTGAGACTTTTCACAGAGGATGTGTGTACCCAGCTACACATCAAAA--- 683
Db 1392 LeuGlnGlnGlnLeuGluAspGluAla-----LysArgAsnTyrGluArgLys 1409
QY 684 -----GAAATGGATAAAATTAAGTGGAAAATTAGAAGATTCAACTAGC 725
Db 1410 LeuAlaGluValThrThrGlnMetGlnGluIleLysLysAlaGluGluAspAlaAsp 1429
QY 726 CTATCAAAATCTTGGATCAGTTCATTCTGTGAAAGACGAGGAACTTCAAAAAGAT 785
Db 1430 LeuAlaLysGluLeuGlu-----GluGlyLysLysArgLeuAsnLysAsp 1444
QY 786 CACTGTGTGAACAACCTACAGAAAATAGGAACAAATGGAAGAAAGTTTGTGTACTGAAA 845
Db 1445 -----IleGluAlaLeuGluArgGlnValLysGluLeuIleAlaGlnAsn 1459
QY 846 AAGAACTGTGCAAGCA---AAGAAATAAATCACAGTTAGAGAACCAAAAGTTAA 902
Db 1460 AspArgLeuAspLysSerLysLysIleGlnSerGluLeuGluAlaThrIleGlu 1479
QY 903 TGGGAACAAGACTCTGCAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAAGA 962
Db 1480 LeuGluAlaGln-----ArgThrLysValLeuGluLeuLysGlnLys 1495
QY 963 AATGCCGAT-----ATATTAAATGAAAAATTAGGGAAGAAATTA 1001

```

Db 1496 AsnPheAspLysIleLeuAlaGluLysAlaIleSerGluGlnIleAlaGlnGluArg 1515
 QY 1002 GAAGAATCGAA-----GAGCAGCATAGGAAGAGTTGAGAAACAAACAA 1049
 Db 1516 AspThrAlaGluArgGluAlaArgGluLysThrLysValLeuSerValSerArgGlu 1535
 QY 1050 CTTGAACAGGCTCTC--AGATAACAAGATATAGAA-----TTGAGAGTGTAGAACT 1100
 Db 1536 LeuAspGluAlaPheAspLysIleGluAspLeuGluAsnLysArgLysThrLeuGlnAsn 1555
 QY 1101 AATTGGAATCAGTTTCTCACACTCATGAAATGAAATATCTCTTACATGAAATTCG 1160
 Db 1556 GluLeuAspLeuAlaAsnThrGlnGlyThrAlaAspLysAsnValHisGlu----- 1573
 QY 1161 ATGTTGAAAAAGGAATTCGCATCTAAACTGCAATAGCCACACTGAAACACCAATAC 1220
 Db 1574 ---LeuGluLysAlaLysArgAlaLeuGluSerGlnLeuAlaGluLeuLysAlaGlnAsn 1592
 QY 1221 CAGAAAAAGGAAAT-----AAATCTTTGAGACATTAAGATTTTAAAGAAAGAAAT 1274
 Db 1593 GluGluLeuGluAspAspLeuGlnLeuThrGluAspAlaLysLeuArgLeuGluValAsn 1612
 QY 1275 -----GCTGAATTCAGATGACCTCAAACTGAAAGAGGAATCATTAAT 1319
 Db 1613 MetGlnAlaLeuArgSerGlnPheGluArgAspLeuLeuAlaLysGluGluGlyAlaGlu 1632
 QY 1320 AAAAGGCGATCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAG----- 1367
 Db 1633 GluLysArgGlyLeuValLysGlnLeuArgAspLeuGluThrGluLeuAspGluGlu 1652
 QY 1368 -----AACCAATGCTCAGCTTCTAAATGGAAGGAAACAAAGCAAAAGAAATA 1415
 Db 1653 ArgLysGlnArgThrAlaAlaValAlaSerLysLysLeuGluGlyAspLeuLysGlu 1672
 QY 1416 CTAGAGCAGAAATGAATCACCACCTCTAGACTGGCTTCTGCTGTACAAGACCATGAT 1475
 Db 1673 IleGluThrThrMetGluMetHisAsnLysValLysGluAspAlaLeu-LysHisAlaLys 1692
 QY 1476 CAATTTGTGACATCAAGAAAGGTCAGAACCTGCTTCCACATTCACAGGAGATGCTGT 1535
 Db 1692 slyLeu-----GlnAlaGlnValLysAspAlaLeuArgAspAlaGluGluAlaLysAl 1710
 QY 1536 TTGCAAGAAATGAATGTTGATG-----TGAGTAGTACCGATATATAA 1580
 Db 1710 aAlaLysGluGluLeuGlnAlaLeuSerLysGluAlaAspGlyLysValLysAlaLeuGlu 1730
 QY 1581 CAATGAGTGCTCCAT-----CAACCATCTTCTGAGGCTCAAGGAAATCCANAAGCT 1634
 Db 1730 uAlaGluValLeuGlnLeuThrGluAspLeuAlaSerSerGluArgAlaArgAlaAl 1750
 QY 1635 AAAAATTAATCTCAATATCAGAGAGATGCTTAAGAGAAATATACATGTTTTCAGGAAC 1694
 Db 1750 aGluThrGluArgAspGluLeuAlaGluLysIleAlaAsnAlaAsnLysGlySerLe 1770
 QY 1695 ATGCACAAAGACCAACGTCAGAAACAGTGTCAAATGAAGAGCTGAACACATGTTTC 1754
 Db 1770 uMetIleAsp-GluLysArgArgLeuGluAlaArgIleAlaThrLeuGluGluLeuG 1790
 QY 1755 AAACGAAACAGAT-----NATGTGAACAAACACACTGANCAGCAGGAGCTCTC 1802
 Db 1790 luGluGluGlnSerAsnSerGluValLeuLeuAspArgAlaAlaAlaArgGlnLeuGlnI 1810
 QY 1803 TAGATCAGAAATATTTCACACTCAAGCAAAATATGTGGCTTCAACAGCAATTAGTTC 1862
 Db 1810 leGluGlnLeuThrGluLeu----- 1817
 QY 1863 ATGCACATTAANGAAGCTGACAAACAAAGCAAGATTAACAATTTCATNTCTTGAG 1922
 Db 1818 --AlaAsnGluLysSerAsnSerGlnLys-----A 1827
 QY 1923 AGGAAATATGNCATCATCTTCTTAAAGAGAAAAATGAGGAGATA 1966
 Db 1827 snGluAsnGlyArgAlaLeuLeuGluArgGlnAsnLysGluLeu 1841

RESULT 6

T08880
 NMMA receptor-binding protein yotiao - human
 C:Species: Homo sapiens (man)
 C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C:Accession: T08880
 R:Lin, J.W.; Wyszynski, M.; Madhavan, R.; Sealock, R.; Kim, J.U.; Sheng, M.
 J. Neurosci. 18, 2017-2027, 1998
 A:Title: Yotiao, a novel protein of neuromuscular junction and brain that interacts with
 A:Reference number: Z16511; MUID:98151389; PMID:9482789
 A:Accession: T08880
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1642 <LIN>
 A:Cross-references: EMBL:AF026245; NID:g2623067; PIDN:AAB86384.1; PID:g2623068
 C:Genetics: T08880
 A:Map position: 7q21-22
 C:Keywords: brain; cerebral cortex; coiled coil; neuromuscular junction; skeletal muscle

Alignment Scores:

Pred. No.: 1.75e-09 Length: 1642
 Score: 278.50 Matches: 158
 Percent Similarity: 39.52% Conservativity: 172
 Best Local Similarity: 18.92% Mismatches: 250
 Query Match: 8.18% Indels: 255
 DB: 2 Gaps: 30

US-09-602-362E-15 (1-2030) x T08880 (1-1642)

QY 135 GCCATTGAAATGCAAAAGCTGTCCAAATAAAGCCTTGGAAATCGAAGATGAA----- 188
 Db 112 SerValGluLeuGluSerGluIleSerThrAlaAspAspCysSerSerGluValAsn 131
 QY 189 -----CAAACATTGAGACGACATGAGATA 212
 Db 132 GlyCysSerPheValMetArgThrGlyLysProThrAsnLeuLeuArgGluGluPhe 151
 QY 213 CTCCCATGAGATCCAAACAAAGAGCTATGAAAGAGTCTTGGGATTCTGAGAGTCTC 272
 Db 152 GlyValAspSerTyrSer-----GluGlnGlyAlaGlnAspSerProThrHis 168
 QY 273 TGTGACACTGTTTCCAGAGAGGATGTGTTCACCAAGCTTACACATCAAAAAAGATA 332
 Db 169 LeuGluMetGluSerGlu-----LeuAlaGlyLysGlnHisGluIle 183
 QY 333 GATAAAATAATGGAATAATTAGAGAGTCTCTCGATAATGATGTTTCTCGAAGGCTCC 392
 Db 184 GluGluLeuAsnArgGluLeuGlu----- 192
 QY 393 TGCAGATGAAGTTCTTATTCCAACTAAAGCTTAGAATTTAGGACATCGAACATCTTC 452
 Db 193 -----MetArgValThrTyrGlyThrGluGlyLeuGln-----GlnLeuGlnGluPhe 208
 QY 453 AAAGCAGAGCTCCCGAGAGACCATCTGCTTCGAGAGCTGCCATTGAAATGCAAAAGTCT 512
 Db 209 GluAlaAlaIleLysGlnArg-----AspGlyIleIleThrGlnLeuThr 223
 QY 513 GTTCCAAATAAAGCCTTGAATTTGAAGATGAACAACATTGAGACGACATGATGTTCT 572
 Db 224 AlaAsnLeuGlnGlnAlaArgArgGluLysAspGluThrMetArgGluPheLeuGluLeu 243
 QY 573 CCTTCAGATCAAAACAAAGAGGCTTGAA--GAAAAATCTTGGGATTCTGAGAGTCTC 629
 Db 244 ThrGluGlnSerGlnLysLeuGlnIleGlnPheGlnGlnLeuGlnAlaSerGluThrLeu 263
 QY 630 CGTGAGACTGTT-----TCACAGAGGATGTGTGTGTACCCAG-----GCT 671
 Db 264 ArgAsnSerThrHisSerSerThrAlaAlaAspLeuLeuGlnAlaLysGlnGlnIleLeu 283
 QY 672 ACACATCAAAAGAAATGATGATAAAATAGTGGAAAAATTAGAAGAT----- 716
 Db 284 ThrHisGlnGlnLeuGluGluGluGlnAspHisLeuLeuGluAspTyrGlnLysLysLys 303

| | | | |
|----|------|---|------|
| Qy | 717 | -----TCACTACGCTATCAAAAATCTTGAT-----ACAGTTCAATCTGTGAA | 761 |
| Db | 304 | GlusppheThrMetGlnIleSerPheLeuGlnGlnLysIleYsValTyRGlumetGlu | 323 |
| Qy | 762 | AGACAAGGGAACCTTCAAAAAGATCAC-----TGT | 791 |
| Db | 324 | GlnAspLysLysValGluAenSerAsnLysGluGlnLysGluThrIle | 343 |
| Qy | 792 | GAACAACGTACAGGAAAAATGGACAACAATGAAAGAAGTTTGTGTACTGAAAAAGAAA | 851 |
| Db | 344 | GluGluLeuAsnThrLysIleIleGluGluGlnLysLysThrLeuGluLeuLysAspLys | 363 |
| Qy | 852 | CTGTCAAGCAAGCAAAAGAAATAAATCACAGTTAGAGAACCAAAAGTTAAATCGGAACA | 911 |
| Db | 364 | LeuThrAlaAspLysLeuLeuGlyGluLeuGlnGluGlnIleValGlnLysAsnGln | 393 |
| Qy | 912 | GAGCTCTGCTGTGAGATTGACTTTAAACAAGAGAGAGAGAGAAATGCGCAT | 971 |
| Db | 384 | GluIleYsAsnMetLysLeuGluLeuThrAsnSerLysGlnLysGluAArgGlnSer--- | 402 |
| Qy | 972 | ATATTAATGAAAAAATTAGGGAAGAATTAGGAAGAAATCGAAGAG----- | 1016 |
| Db | 403 | -----SerGluGluIleLysGlnLeuMetGlyThrValGluGluLeuLysArgAsn | 420 |
| Qy | 1016 | ----- | 1016 |
| Db | 421 | HisLysAspSerGlnPheGluThrAspIleValGlnArgMetGluGlnGluThrGlnArg | 440 |
| Qy | 1016 | ----- | 1016 |
| Db | 441 | LysLeuGluGlnLeuArgAlaGluLeuAspGluMetTyRglyGlnGlnIleValGlnMet | 460 |
| Qy | 1017 | -----CAGCATPAGAAAGAGTTAGAACTG-----AAA | 1043 |
| Db | 461 | LysGlnGluLeuIleArgGlnHisMetAlaGlnMetGluGluMetLysThrArgHisLys | 480 |
| Qy | 1044 | CAACAACCTGAACAGGCTCTCAG----- | 1067 |
| Db | 481 | GlyGluMetGluAsnAlaLeuAArgSerTySerAsnIleThrValAsnGluAspGlnIle | 500 |
| Qy | 1068 | -----ATACAAGATATAGAATTGAAGAGTGTAGAAAGTAATTTGAAT | 1109 |
| Db | 501 | LysLeuMetAsnValAlaIleAsnGluLeuAsnIleLysLeuGlnAspThrAsn---Ser | 519 |
| Qy | 1110 | CAGTTTCTCACACTCATGAATAGAAAATATCTTTACATGAAAATTCATGTTGAA | 1169 |
| Db | 520 | GlnLysGluLysLeuLysGluGluLeuGlyLeuIleLeuGluLysCysAlaLeuGln | 539 |
| Qy | 1170 | AAGAAAATGCGATCTTAAAACTGGAATAGCCACACTGAAACCAAAATACCAAG----- | 1223 |
| Db | 540 | ArgGlnLeuGluAspLeuValGluGluLeuSerPheSerArgGluGlnIleGlnArgAla | 559 |
| Qy | 1224 | -----GAAAAGAAAATAAATAC----- | 1241 |
| Db | 560 | ArgGlnThrIleAlaGluGlnGluSerLysLysLeuAsnGluAlaHisLysSerLeuSerThr | 579 |
| Qy | 1242 | TTTCAGGACATTAAGATTTTAAAAAGAAAAGAAATGCTGAAGTTCAGATGACCCATAAATCG | 1301 |
| Db | 580 | ValGluAspLeuLysAlaGluIleValSerAlaSerGluSerArgLysGluLeuGluLeu | 599 |
| Qy | 1302 | AAAGAGAAATCAATTAACAAAGGCGATCTCAATATAGTGGGAGCTTAAAGTTCGATA | 1361 |
| Db | 600 | LysHisGluAla-----GluValThrAsnTyrlsIleLysLeuGluMetLeuGlu | 616 |
| Qy | 1362 | GCTGAGAACACAATGCTCACTTCTTAATTCGAAGAAAAACAAGCAAGAAATA----- | 1415 |
| Db | 617 | LysGluLysAsnAlaValLeuAspArgMetAlaGluSerGlnGluAlaGluLeuGluArg | 636 |
| Qy | 1416 | -----CTAGAGGCAGAAAT | 1430 |
| Db | 637 | LeuArgThrGlnLeuLysPheSerHisGluGluGluSerLysLysLysGluAspLeu | 656 |

| | | | |
|----|------|---|------|
| Qy | 1431 | GAATCACACATCTCTAGACTGGCTTCGTGTACAGACCATGATCAATTTGCACATCA | 1490 |
| Db | 657 | GlulGluHisArgIleAsnIleGluLysLeuLysAspAsn-LeuGlyIleHisTyrLy | 676 |
| Qy | 1491 | AGAAAGACTCAAGAACCTG-----CTTTCACATTCGAG----- | 1524 |
| Db | 676 | sGlnGlnIleAspGlyLeuGlnAsnGluMetSerGlnLysIleGluThrMetGlnPheGl | 696 |
| Qy | 1525 | -----GAGATCCTGTTTGCAAGAAAAATGAATGTGATGTAGTAGTACCGATATATAA | 1580 |
| Db | 696 | uLysAspAsnLeuIleThrLysGlnAsnGlnLeuIleLeuGluIleSerLysLeuLysAs | 716 |
| Qy | 1581 | CAATGAGTGCTCCATCAACCATCTTCTGAAGCTCAAGAGAAATCCANAAGCCATAAAAT | 1640 |
| Db | 716 | p-----LeuGlnGlnSerLeuValAsnSerLysSerGluGluMetThrLeuGlnIle | 733 |
| Qy | 1641 | TAAT--CTCAATATGCAGGAGATGCTCTAAGAGAAAAATACATTTGGTTTCAGACACATG | 1697 |
| Db | 733 | eAsnGluLeuGlnLysGluIleGluIleLeuArgGlnGluLysGluLysGlyThr-L | 753 |
| Qy | 1698 | CACAAAGACACCAACGTGAACACAGTGT-----CAAATCAAGG | 1736 |
| Db | 753 | euGlnGlnGluValGlnGlnLeuGlnLeuLysThrGluLeuLeuGluLysGlnMetLysG | 773 |
| Qy | 1737 | AAGCTGAACATGTTNCCAANGCAA----- | 1762 |
| Db | 773 | lulysGluAsnAspLeuGlnGluLysPheAlaGlnLeuGluAlaGluAsnSerIleLeuL | 793 |
| Qy | 1763 | -----CAAGATNATGTGAACAAACACACT-----GANCACGAGG | 1796 |
| Db | 793 | ysAspGluLysLysThrLeuGluAspMetLeuLysIleHisThrProValSerGlnGluG | 813 |
| Qy | 1797 | AGTCTCTA-----GATCAGA | 1811 |
| Db | 813 | luArgLeuIlePheLeuAspSerIleLysSerLysSerLysAspSerValTrpGluLysG | 833 |
| Qy | 1812 | AATATTATTCACATAAAGCAAAATATGCGCTTCAACAGCAATTAGTTTCATGCACATA | 1871 |
| Db | 833 | luIleGluIleLeuIleGluGluAsnGluAspLysGlnGlnCysIleGlnLeuAsnG | 853 |
| Qy | 1872 | ANGAAAGCTGACACAAAGCAAGATAACAATTTGATNTTTCATTNCTTGAGAGAAAAATG | 1931 |
| Db | 853 | luGluIleGluLysGlnArgAsnThr-PheSerPheAlaGluLysAsnPheGluValAsnT | 873 |
| Qy | 1932 | CNCATCATCTTCTAAAGAG----- | 1951 |
| Db | 873 | yrGlnGluLeuGlnGluGluTyrAlaCysLeuLeuLysValLysAspAspLeuGluAspS | 893 |
| Qy | 1952 | --AAAAATGAGGAGATATTNTTATTCATNATAACCACTTTAAAA | 1990 |
| Db | 893 | erLysAsnLysGlnGluLeuGluTyrLysSerLysLysLys | 906 |

RESULT 7

T13030.
 microtubule binding protein D-CLIP-190 - fruit fly (*Drosophila melanogaster*)
 C:Species: *Drosophila melanogaster*
 C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C:Accession: T13030
 R:Lantz, V.A.; Miller, K.G.
 J. Cell Biol. 140, 897-910, 1998
 A:Title: A class VI unconventional myosin is associated with a homologue of
 A:Reference number: Z17598; MUID:98139549; PMID:9472041
 A:Accession: T13030
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-1690 <LAN>
 A:Cross-references: EMBL:AF041382; NID:G2773362; PID:G2773363; PIDN:AAB96783
 A:Experimental source: strain Oregon R
 C:Genetics:
 A:Cross-references: FlyBase:FBgn0020503
 C:Keywords: cytoskeleton

 Alignment Scores:

Pred. No.: 2,66e-09 Length: 1690
 Score: 275.50 Matches: 174
 Percent Similarity: 37.96% Conservative: 127
 Best Local Similarity: 21.94% Mismatches: 283
 Query Match: 8.10% Indels: 229
 DB: 2 Gaps: 35

US-09-602-362E-15 (1-2030) x T13030 (1-1690)

QY 39 GGAATGAAAGTTCTTATTCACAAAGCCCTAGAAATGATGAGCATG-----CAAACT 92
 Db 330 GYThrMetAsnSerIleAlaThrAlaThrSerArgMetArgMetAsnAlaGlnGln 349
 QY 93 TTCAAAGCAGCGCTCCCGAAGAGCATCTGCTCGAGCGCTGCCATTGAAATG----- 146
 Db 350 ArgLysSerSerThrProValLysProIleLeuAlaThrProLysSerGlnPheSerMet 369
 QY 146 ----- 146
 Db 370 GlnAspLeuLeuArgGluLysGlnHisValGluLysLeuMetValGluArgAspLeu 389
 QY 147 ----CAAAGTCTCTTCCAAATAAGCCCTGGAATTG---AAGAATGAACAAACATTGAGA 200
 Db 390 AspArgGluAspAlaGlnHisGlnAlaLeuGlnLeuGlnLysAsnIleAsnGluLeuLys 409
 QY 201 GCA-----GATGAGATACCTCCATCAGAAATCCAAACAAAGAGCATATGAA 245
 Db 410 AlaArgIleValGluLeuGluSerAlaLeuGluArgLysLysThrGluGluLeu 429
 QY 246 GAAAGTCTTGGATCTGAGACTCTCTGT-----GAGACTGTTTCAGAAAGCATGTG 299
 Db 430 GlnCysSerIleAspGluAlaGlnPheCysGlyAspGluLeuAsnAlaGlnSerGlnVal 449
 QY 300 TGTTTACCAAGGCTACACATCAAAAGAAATAGATAAAATAATGGAATA----- 353
 Db 450 -----TyrLysGluLysIleHisAspLeuGluSerLysIleThrLys 463
 QY 354 -----GAGAGTCTCTGATATGAT 374
 Db 464 LeuValSerAlaThrProSerLeuGlnSerIleLeuProProAspLeuProSerAspAsp 483
 QY 375 GGTTCCTGAAG-----GCTCCCTGCAGAAATGAAAGTTTCTTCCAACTAAAGCC 425
 Db 484 GlyAlaLeuGlnGluLeuAlaGlnLeuGlnGluLysMetThrIleGlnGlnLysGlu 503
 QY 426 TTGAATTTGATG----- 437
 Db 504 ValGluSerArgIleAlaGluGlnLeuGluGluGlnArgLeuArgGluAsnValLys 523
 QY 438 -----GACATGCAAACTTTCAAAGCAGAGCCCTCCCGAAGGCATCTGCCTTC 485
 Db 524 TyrLeuAsnGluGlnIleAlaThrLeuGlnSerGluLeuValSerLysAspGluAlaLeu 543
 QY 486 GAGCCTGCCAATGAAATGCAAGTCTGTTCCAAATGAAAGCCCTGGAATG----- 536
 Db 544 GluLysPheSerLeuSerGluCysGlyIleGluAsnLeuArgArgGluLeuAlaLeuLeu 563
 QY 537 -----AAGAAATGAA---CAAAATTTGAGAGCAGATCAGATGTTCCCTTCAGAAATCAAAA 587
 Db 564 LysGluGluAsnGluLysGlnAlaGlnGluAlaGlnAlaGluPheThrArgLysLeuAla 583
 QY 588 CAAAGAAGGTTGAAGAAATTTCTGGGATCTGAG----- 623
 Db 584 GluLysSerValGluValLeuArgLeuSerSerGluLeuGlnAsnLeuLysAlaThrSer 603
 QY 624 ---AGTCTCGT---GAGACTGTTTCACAGAGGATGTGTGTGATCCCAAGGCTACACAT 677
 Db 604 AspSerLeuGluSerGluArgValAsnLysSerAspGluCysGluIleLeuGlnThrGlu 623
 QY 678 -----CAAAAAAATGATGATAAATAAGTGA 704
 Db 624 ValArgMetArgAspGluGlnIleArgGluLeuAsnGlnGlnLeuAspGluValThrThr 643

QY 705 AAATTAGAA-----GATTCACACTAGCTATCAAAATCTTGGATACAGTTCA 752
 Db 644 GlnLeuAsnValGlnLysAlaAspSerSerAlaLeuAspMetLeu----- 659
 QY 753 TCTTGTGAAGACAGGGAACCTTCAAAAGATCACTGTGGAACAAGTACAGGAAATG 812
 Db 660 -----ArgLeuGlnLysGluGlyThrGluGluLysSerThrLeuLeu 673
 QY 813 GAACAAAATGAAAGAAAGTTTGTGTACTGAAAGAAACTGTCTCAGAACAAAGAAATA 872
 Db 674 GluLysThrGluLysGluLeuValGlnIleLysGluGlnAlaAlaLysThrLeuGlnAsp 693
 QY 873 AATTCAGTTAGAACCAAAAGTAAATGGACACAGAGCTCTGCAGTGTGAGATTG 932
 Db 694 LysGluGlnLysGlnLysGln-----IleSerAspLeuLysGln 706
 QY 933 ACTTTAAACCAAGAAAGAGAGAGAAATGCGCATATATTAATGAAATAATAG 992
 Db 707 LeuAlaGluGlnLysLeuValArgGluLysThrGluAsnAlaIleAsnGlnIleGln 726
 QY 993 GAGAAATTAGGAAGAAATCGAAGAGAG-----CATAGGAAGAGTTAGAA----- 1037
 Db 727 LeuGluLysGluSerIleGluGlnGlnLeuAlaLeuLysGlnAsnGluLeuGluAspPhe 746
 QY 1038 GTGAACAACAACCTTGAAACAGGCTCTCAGAATACAAAGATATAGAA----- 1082
 Db 747 GlnLysLysGlnSerGluSerGluValHisLeuGlnGluIleLysAlaGlnAsnThrGln 766
 QY 1083 -----TTGAAGAGTGTAGAAAGTAATTTGAA 1109
 Db 767 LysAspLeuGluLeuValGluSerGlyGluSerLeuLysLeuGlnGlnLeuGlu 786
 QY 1110 CAGGTTTCACACTCATGAAATGAAATATCTCTTACATGAAATTCATGTTGAAA 1169
 Db 787 GluLysThrLeuGlyHisGluLysLeuGlnAlaAlaLeuGlu-----LeuLys 803
 QY 1170 AAGGAA-----ATTGCCATCTGAAACTGAAATAGCCACACTGAAACACCAA 1217
 Db 804 LysGluLysGluThrIleLysGluLysGluGlnGluLeuGlnGlnLeuGlnSerLys 823
 QY 1218 TACCAGAAAGAAAT----- 1235
 Db 824 SerAlaGluSerGluSerAlaLeuLysValValGlnValGlnLeuGluGlnGln 843
 QY 1236 -----AAATCTTTGAGGACAT 1253
 Db 844 GlnAlaAlaAlaSerGlyGluGluGlySerLysThrValAlaLysLeuHisAspGluIle 863
 QY 1254 AAGATTTTAAAGAAAGAAATGCTGAACTTCAGATGACCTTAAACTGAAAGAGGATCA 1313
 Db 864 SerGlnLeuLysSerGlnAlaGluGluThrGlnSerGluLeuLysSerThrGluSerAsn 883
 QY 1314 TTAATAAAAGGCGATCTCAATAT-----AGTGGCGAGCTTAAAGTTCTTGATAGCT 1364
 Db 884 LeuGluAlaLysSerLysGlnLeuGluAlaAlaAsnGlySerLeuGlu----- 899
 QY 1365 GAGAACCAATGCTCACTCTCTAAATTTGAAGAAAAACAAGACAAAGAAATCTAGAGGCA 1424
 Db 900 GluGluAlaLysLysSerGlyGlnLeuGlnGluGlnIleThrLys-----LeuLysSer 917
 QY 1425 GAAATTCAAATCACACCATCTCAGCTGCTGCTGTACAGACCATGATCAATTTGTG 1484
 Db 918 ValGluGluThrGlnAlaAlaLeuSerSerTyr-----HisThrAspValGlu 934
 QY 1485 ACATCAAGAAAAAGTCAAGAACCTGCTTCCCATTCGAGGAGATGCTGTTTGCAAAGA 1544
 Db 935 SerLysThrLysGlnLeuGlu-----AlaAlaAsnAlaAlaLeuGlu--- 948
 QY 1545 AAAATGATGATTTGAT-----GTGAGTAGTAGCCGATATATAACAAATGAGGT 1589
 Db 949 LysValAsnLysGluTyrAlaGluSerArgAlaGluAlaSerAspLeu-GlnAspLysVa 968
 QY 1590 G-----CTCCATCAACCACTTTCTGAAAGCTCAAGGAAATCCANAAG 1631

Db 968 llysgluilthrThrLeuHisAlaGluAlaGluArgSerSerSerAl 988
 QY 1632 CTTAAATTAATCTCAATATGCGAGGATGCTCTAAGAGAAATACAT 1681
 Db 989 aleuHisThrLysLeuSerLysPheSerAspGluileAlaThrGly-HisLysGluLeu 1008
 QY 1682 -----TGGTTTCAGGAACATGCACAAAGACCAACGCTGAACACACAGT 1724
 Db 1008 hr-SerLysAlaAspAlaTrpSerGlnGluMetLeuGlnLysGluLys 1023
 QY 1725 GTCAATGAGGAAGCTGACACATGTNTCAANGCAACAGATNATGTGACAAACACA 1784
 Db 1024 --GluLeuGlnGluLeuArgGlnGlnLeuGlnAspSerGlnAspSerGlnThrLysLeu 1043
 QY 1785 CTGANCAGCAGGAGTCTCTAGATCAGAAATATTTCAACTACAAAGCAAAATATGTGGC 1844
 Db 1043 ysAlaGluGlyGluArg--LysGluLysSerPheGluGluSerIleLysAsn 1060
 QY 1845 TTCAACAGCAATAGTTTCATGCGACATAANGAAAGC 1879
 Db 1060 euGlnGluGluValThrLysAlaLysThrGluAsn 1071
 RESULT 8
 C71622
 hypothetical protein PF0145c - malaria parasite (Plasmodium falciparum)
 C/Species: Plasmodium falciparum
 C/Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 C/Accession: C71622
 R/Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 S.; Partea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
 Science 282, 1126-1132, 1998
 A/Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A/Reference number: A71600; MUID:99021743; PMID:9804551
 A/Accession: C71622
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-1979 <GAR>
 A/Cross-references: GB:AE001375; GB:AE001362; NID:93845105; PIDN:AC71819.1; PID:9384510
 A/Experimental source: clone 3D7
 C/Genetics:
 A/Gene: PF0145c
 Alignment Scores:
 Pred. No.: 8,66e-09 Length: 1979
 Scores: 267.00 Matches: 185
 Percent Similarity: 37.06% Conservative: 130
 Best Local Similarity: 21.76% Mismatches: 268
 Query Match: 7.85% Indels: 268
 DB: 2 Gaps: 38
 US-09-602-362E-15 (1-2030) x C71622 (1-1979)
 QY 141 GAATGCAAAAGCTGTGTTCCAAATAAAGCCTTGAATGAAGATGAA-----CAAAACA 194
 Db 118 GlutLeuSerLysAsnIleHisLysAspAsnGluLeuGluAsnGlnLeuLysAspThr 137
 QY 195 TTGAGA-----GCAGATGAGATATCTCCCATCAGAAATCCAA----- 230
 Db 138 LeuLysSerIleSerSerLeuSerAsnLysIleValAsnThrGluSerLysIleGluGlu 157
 QY 231 -----CAAAAGGACTATGAAGAA 248
 Db 158 LeuGluLysGluLeuLysAspLysAsnIleAspAsnAspTyrGluAsn 177
 QY 249 AGTCTTGGGATCTGAGAGTCTCTGTCAGACTGTTTCACAGAG-----GATGTGTGTTA 305
 Db 178 LysLeuLysGluLysGlu-----AspPheValLysGlnLysIleAspMetLeuAsn 194
 QY 306 CCCAAGGCTACA-----CATCAAAAGAAATAGAT-----AAAATA 341
 Db 195 GluLysGluAsnLeuLeuGlnGluLysGluLeuAspIleAsnLysArgGluLysLysIle 214

QY 342 AATCGAAA----- 350
 Db 215 AsnGluLysGluLysAsnIleIleLysLysGluGluThrPheHisAsnIleGluLysGlu 234
 QY 351 ---TTAAGAGAGTCTCCTGATAATGATGTTTCTGAAGGCTCCCTGCAGATGAAGTT 407
 Db 235 TyrLeuGluLysAsnLysGluArgGlu----- 243
 QY 408 TCTATTCCAACTAAAGCCTTAGAATTCATGAC-----ATGCAAACTTCAAA 455
 Db 244 -----ThrIleSerIleGluIleIleAspIleLysLysHisLeuGluLysLeuLys 260
 QY 456 GCAGAGCCTCCGAGAGCCATCTGCTTCGAGCCTGCCATTCGAATGCAAAATCTGTGT 515
 Db 261 IleGluIleLysGluLysLysGluAspLeuGluAsnLeuAsnLysLysLeuLeuSerLys 280
 QY 516 CCAAAATAAGCCTTGGAAATTGAAG-----AATGAACAACAATTCAGAGCA 560
 Db 281 GluAsnValLeuLysGluLeuLysGlyCysValLysGluLysAsnGluThrIleAsnSer 300
 QY 561 -----GATCAGATGTTCCCTTCAGAAATCAAAACAAAGAGGTTGAA----- 602
 Db 301 LeuAsnAspAsnIleIleGluLysGluLysLysTyrLysLeuLeuGluTyrGluLeuGlu 320
 QY 603 -----GAAATTCCTGGGATTCAGAGT 626
 Db 321 GluLysAsnLysGlnIleAspLeuLeuAsnLysGlnGluLysGluLysGluLys 340
 QY 627 CTCGCTGAGACTCTTTCACAGAGAGGATGTGTGTACCCAGAGCTACACATCAAAAGAA 686
 Db 341 GluArgGluLysGluLysGluArgGlu-----LysGluLysGluLysGlu 355
 QY 687 ATCGATAAATAAGTGGAAATTAGAAGATCAACTAGCCTATCAAAAATCTTGGATACA 746
 Db 356 TyrAspThrLeuIleLysGluLeuLysAspGluLys-----IleSerIleLeuGluLys 373
 QY 747 GTTCATCTCTGTCAA---AGAGCAAGGGAACITCAAAAGAT-----CAC 788
 Db 374 ValHisSerIleLysValArgGluMetAspIleGluLysArgGluHisAsnPheLeuHis 393
 QY 789 TGTGAACACGTCACAGAAAATGGAACAAATCAAAAGAGAGTTTGT----- 836
 Db 394 Met-GluAsp-----GlnLeuLysAspLeuLysAsnSerPheValLysAsnAsn 410
 QY 837 GTACTGAAAAGAAAGTGTGACAGCAAAAGAAATAAATCAAGTTAGAGAACCAAAA 896
 Db 411 GlnLeuLysValTyrLysCysGluIleLysAsnLeuLysThrGluLeuGlu----- 427
 QY 897 GTTAAATGGCAACAAGAGCTCTGC-----AGTGTGAGATTG 932
 Db 428 ---LysLysGluLysGluLeuLysAspIleGluAsnValSerLysGluIleAsnLys 446
 QY 933 ACTTTAAACAGAGAAGAGAGAGAGAAATGCGATATATTAATGAAAAAATTAGG 992
 Db 447 LeuIleAsnGlnLeuAsnGluLysGluLysGlnIleLeuAlaPheAsnLysHisLys 466
 QY 993 GAAGAA-----TTAGGAAGATCGAA 1013
 Db 467 GluGluIleHisGlyLeuLysGluGluLeuLysGluSerValLysIleThrLysIleGlu 486
 QY 1014 GAGCAGCATAGGAAGAGTTA---GAAGTGAACAA-----CAACTGAACAG 1058
 Db 487 ThrGlnGluLeuGlnGluMetValAspIleLysGlnLysGluLeuAspGlnLeuGlnGlu 506
 QY 1059 GCTCTCAGAAATACA-----GATATGAATTCGAAGAGTGTGAAGATTAATTG 1106
 Db 507 LysTyrAsnAlaGlnIleGluSerIleSerIleGluLeuSerLysLysGluLysGlu 526
 QY 1107 AATCAGGTTTTCACACTCATGAAAATGAA----- 1136
 Db 527 AsnGlnTyrAsnThrThrIleGluGluIleAsnAsnLeuAsnGluLysLeuGluGlu 546
 QY 1137 -----AATTATCTCTTA 1148

| | | | | |
|----|------|---|------|--|
| Db | 547 | ThrasnLysGluTyrThrasnLeuGlnAsnAsnTyrThrasnGluIleAsnMetLeuAsn | 566 | |
| Qy | 1149 | CATGAATTCATCTGAAAGGAAATGCCATGCTAAACATGGAATAGCCACACTG | 1208 | |
| Db | 567 | AsnAspIleHisMetLeuAsnGlyAsnIleLysThrMetAsnThrGlnIleSerThrLeu | 586 | |
| Qy | 1209 | AAA-----CACCAATACCAGGAAAGGAAAAATAAATCTTTGAGCATTAAAGATT | 1259 | |
| Db | 587 | LysAsnAspValHisLeuLeuAsnGluGlnIleAspLysLeuAsnAsnGluLysGlyThr | 606 | |
| Qy | 1260 | TTAAAGAAAAAGATCTCACTTCTTAAATTTGAAGGAAAAACAAGAGGAA | 1310 | |
| Db | 607 | LeuAsnSerLysIleSerGluLeuAsnValGlnIleMetAspLeuLysGluGluLysAsp | 626 | |
| Qy | 1311 | TCATTAACTAAAGGCGCATCTCAATATAGTGGCGAGCTTAAAGTTCTGATA----- | 1361 | |
| Db | 627 | PheLeuAsnAsnGlnIleValAspLeuSerAsnGlnIleAspLeuLeuThrArgLysMet | 646 | |
| Qy | 1362 | -----GCTGAGACACAATGCTCTTCTTAAATTTGAAGGAAAAACAAGAGGAA | 1415 | |
| Db | 647 | GluGluLysGluAsnLysMetLeuGluGlnGluAsnLysTyrLysGlnGluMetGluLeu | 666 | |
| Qy | 1416 | CTAGAGGCAGAAATT----- | 1430 | |
| Db | 667 | LeuArgGlyAsnIleLysSerSerGluAsnIleLeuAsnAsnAspGluGluValCysAsp | 686 | |
| Qy | 1431 | -----GATCACCACCAT | 1442 | |
| Db | 687 | LeuLysArgLysLeuSerLeuLysGluSerGluMetLysMetMetLysGluGluHisAsp | 706 | |
| Qy | 1443 | CCTAGACTGGCTCTGCTCTACAGACCATGATCAATTTGTGACATCAA----- | 1491 | |
| Db | 707 | LysLysLeuAlaGluLeuLysAspCysAspValArgIle-ArgGluMetAsnGluLys | 726 | |
| Qy | 1492 | -----GAAAAAGTCAAGAACCTGCTTTCCACATTGCA | 1523 | |
| Db | 726 | sAsnGluAspLysIleAsnMetLeuLysGluGluTyrGluAspLysIleAsnThrLeuLys | 746 | |
| Qy | 1524 | GGAG-----ATGCTGTTTGGCAAGAAAAATCAATGTTGATGTGAGTAGT | 1568 | |
| Db | 746 | sGluGlnAsnGluAspLysIleAsnThrLeuLysGluGln----- | 759 | |
| Qy | 1569 | ACCGATATATAACAATGAGGTGCTCCATCAACCATCTTTCTGAAGCTCAAGGAAATCCAN | 1628 | |
| Db | 760 | -----AsnGluAspLysIleAsnThrLeuLysGluGluTyrGluHisLysIleAs | 776 | |
| Qy | 1629 | AGCCTAAAAATTAATCTCAATATGCGAGAGATGCTCTAAGAGAAATACATTGGTTTC | 1688 | |
| Db | 776 | nThrMetLysGluGluTyrGluHisLysIleAsnThrLeuAsnGluGln-AsnGluHisL | 796 | |
| Qy | 1689 | AGGAACATGCACAAAGACCAACGAGGAAACACAGTGTCAA---ATGAAGGAA----- | 1739 | |
| Db | 796 | ysIleAsnThrLeuAsnGluGlnAsnGluHisLysIleAsnThrMetLysGluGluTyrG | 816 | |
| Qy | 1740 | CTGAACACATGTTTCAAAAGCAACAGATATGTGAACAAACACACTGANCAGCAGAGT | 1799 | |
| Db | 816 | luAspLysMetAsnThrLeuAsnGluGlnAsnGluAspLysMetAsnSerLeuLysGluG | 836 | |
| Qy | 1800 | CTCTAGATCAGAAATTTTCACTACAAAGCAAAATATATGTGGCTTCAACAGCAATTAG | 1859 | |
| Db | 836 | luTyrGluAsnLysIleAsnGlnIleAsnSerAsnAsnGluIleLysIleLysAspValV | 856 | |
| Qy | 1860 | TTTCATGCACATAAAGAGCTGCAACAAAGCAAGATTAACATTAATTCATTTCTTT | 1919 | |
| Db | 856 | alaAsnGluTyrIleGluGlu-ValAspLysLeuLysValThrLeuAsp----- | 871 | |
| Qy | 1920 | GAGAGAAAT-----GNCATCATCTTCTTAAAGAGAAA | 1954 | |
| Db | 872 | GluLysLysGlnPheAspLysGluIleAsnTyrAlaHisIleLysAlaHisGluLys | 891 | |
| Qy | 1955 | AATGAGGAGATATTTNATTACNATAACCATTTAAAA-----AACCCGTATATT | 2002 | |

892 GluGlnIleLeuLeuThrGluMetGluGluLeuLysCysGlnArgAspAsnLysTyrSer

2003 TCAATATGGAATAAAAAAANA 2026

912 AspLeuTyrGluLysTyrIleLys 919

RESULT 9

T14156

kinesin-related protein - African clawed frog

C;Species: Xenopus laevis (African clawed frog)

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000

C;Accession: T14156

R;Wood, K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.

Cell 91, 357-366, 1997

A;Title: CENP-B is a plus end-directed kinetochore motor required for metaphase chromosome

A;Reference number: Z17893; MUID:98028574; PMID:9363944

A;Accession: T14156

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-2954 <WOO>

A;Cross-references: EMBL:AF027728; NID:g2586070; PID:g2586071; PIDN:AAC60300.1

C;Genetics:

A;Gene: XCENP-B

C;Superfamily: centromere protein E; kinesin motor domain homology

Alignment Scores: 8.75e-09 Length: 2954

Pred. No.: 267.00 Matches: 108

Score: 44.00% Conservative: 112

Percent Similarity: 44.00% Mismatches: 180

Best Local Similarity: 21.60% Indels: 100

Query Match: 7.85%

DB: 2 Gaps: 15

US-09-602-362B-15 (1-2030) x T14156 (1-2954)

Qy 12 AAAGATGCTTCTGAAGCTAACTCGGAATGAAAGTTTCTATTCCAACTAAAGCTTA 71

Db 1874 ArgAspAspLeuAsnGlnSerLysGlnHisLeuPheSerGluIleGluThrLeuSerLeu 1893

Qy 72 GAATTGATGACATGCAAACTTTCAAAGCAGACGCTCCCGAG-----AAGCATCTGCC 125

Db 1894 SerLeuLysGluLysGlu---PheAlaLeuGluGlnAlaGluLysAspLysAlaAspAla 1912

Qy 126 TTGAGCTCCATTTGAATGCAAAAGTCTGTTCCTCAATTAAGCTTGGAAATGAGAAT 185

Db 1913 AlaArgLysThrIleAspIleThrGluLysIleSerAsnIleGluGlnLeuGln 1932

Qy 186 GAACAAACATGAGACAGATGAGATACCTCCCATCAGAAATCCAAACAAAGGACTATGA 245

Db 1933 GlnAlaThrAsnLeuLysGluThrLeuTyrGluArgGluSerLeuIleGlnCysLysGlu 1952

Qy 246 GAAAGTTCTTGGATCTTGAGCTCTCTGTGAGACTGTTTCACAGAGATGCTGTTT 305

Db 1953 GlnLeuAlaLeuAsnThrGluHisLeuArgGluThrLeuLysSerLysAspLeuAlaLeu 1972

Qy 306 CCCAAGCTACACATCAAAAGAA-----ATAGATAAA 338

Db 1973 GlyLysMetGluGlnGluArgAspGluAlaAlaAsnLysValIleAlaLeuThrGluLys 1992

Qy 339 ATAAATCGAAAAATTAGAGAGTCTCTGATATATGATGCTTTCTGAAGGCTCCCTCCAGA 398

Db 1993 MetSerSerLeuGluGlnGlnIleAsnGluAsnValThrThrLeuLysGluGlyGlu 2012

Qy 399 ATGAAATTTCTATCCCACTAAAGCTTAGAATTGATGACATGCAAACTTTCAAAGCA 458

Db 2013 GluLys-----GluThrPheTyrLeu 2019

Qy 459 GAGCTCCCGAGAGCCCATCTGCTTCGACCTGCCATTTGAATGCAAAAGTCTGTTC 518

Db 2020 GlnArgProSerLysGlnGlnSerSerGlnMetGluGluLeuArgGluSerLeuLys 2039

Qy 519 AATTAAGCCCTTGGAAATGAAGATGAACAAACATTGAGACAGATCAGATGTTCCCTTCA 578

```
Db      2040 ThrLysAspLeuGlnLeu----- 2045
QY      579 GAATCAAAACAAAGAAGGTTGAAGAAAATTTCTGGATTCTGAG----- 623
Db      2046 GluGluAlaGluLysGluLeuSerGluAlaThrAsnGluLeuLysAsnLeuThrAlaLys 2065
QY      624 -----AGTCTCGTGAGACTTTTCACAGAAGGATGTGTGTACCAAGGCTACACAT 677
Db      2066 IleSerSerLeuGluGluGluLeuGlnAsnAlaSerIleLeuAsnGluAlaValSer 2085
QY      678 CAAAAGAAGATGGATAAATAGTGGAAAATAGAAATTCACCTAGCTATCAAAAATC 737
Db      2086 GluArgGluAsnLeuArgHisSer---LysGlnGlnLeuValSerGluLeuGluGlnLeu 2104
QY      738 TTGGATACAGTTCATCTTGTGAAGAGCAAGGAACTTCAAAAAGATCAGCTGTGAACA 797
Db      2105 SerLeuThrLysSerArgAspHisAlaPheAlaGlnSerLysArgGluLysAspGlu 2124
QY      798 CGTACAGGAAAATGGAACAAATGAAAAGAAAGATTTGTGTACTCAAAAAGAACTGTCA 857
Db      2125 AlaValAsnLysIleAlaSerLeuAlaGluGluLysIleLeuThrLysGluMetAsp 2144
QY      858 GAAGCAAAAGAAATAAATTCAGTTAGAGAACCAAAAAGTTAAATGGAAACAAGAGCTC 917
Db      2145 GluPheArgAspSerLysGluSerLeuGlnGlnSerSerHisLeuSerGluGluLeu 2164
QY      918 TGCAGTGTGAGATTGACTTTTAAACCAAGAGAAGAGAGAGAAATGCCGATATATTA 977
Db      2165 CysThrTyThrGluLeuGluMetLysGlnMetLysGlnGlnLysGluAsp-----Ile 2181
QY      978 AATGAAAATATAGGGAAGAAATAGGAAGAAATCGAAGAG-----CAGCATAGGAAAG 1031
Db      2182 AsnAsnLysLeuAlaGluLysValLysGluValAspGluLeuGlnHis----- 2198
QY      1032 TTAGAGTGAACAACACTTGAACAGGCTCTCAGAAATACAGATATAGAAATTGAAGAGT 1091
Db      2199 -----LeuSerSer 2201
QY      1092 GTAGAAGTAATTTGATCAGGTTTCTCACACTCATGAAATGAAATATTCTCTTACAT 1151
Db      2202 LeuLysGluGlnLeuAspGlnIleGlnMetGluLeuArgAsnGluLysLeuArgAsnTy 2221
QY      1152 GAAATATGC-----ATGTTGAAAAGGAAATGCGATCTGCTAAAACCTGGAATA 1199
Db      2222 GluLeuCysGluLysMetAspIleMetGluLysGluIleSerValLeuArgLeu----- 2239
QY      1200 GCCACACTGAAACACCAATACAGGAAAGGAAATAAATACTTTTGAGGACATTAAAGATT 1259
Db      2240 -----MetGlnAsnGluProGlnGlnGluLysAspValAlaGluArgMetAspIle 2257
QY      1260 TTAAGAAGAAAGAAATGCTGAATTCAGATGACCCCTAAAACCTGAAAGAGCAATATCACT 1319
Db      2258 LeuGluSerArgAsnGlnGluIleGln-----GluLeuMetGlu 2270
QY      1320 AAAAGGGCATCTCAATATAGTGGGAGCTTAAAGTTCTGTAGCTGAGAAACACATGCTC 1379
Db      2271 LysIleSerAlaValTySerGluGln-----HisThrLeuLeu 2283
QY      1380 ACTTCTAAATTGAAGGAAAACAGACAAAGAAATACTAGAGCGAGAAATTCGAATCACAC 1439
Db      2284 SerSerLeuSerSerGlu-----LeuGlnLysGluThrGluAlaHis 2297

RESULT 10
S28261
centromere protein E - human
N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E
C:Species: Homo sapiens (man)
C>Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C:Accession: S28261
Rifen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
Nature 359, 536-539, 1992
A:Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
A:Reference number: S28261; MUID:93024922; PMID:1406971
```

```
A:Accession: S28261
A:Molecule type: mRNA
A:Residues: 1-2663 <YEN>
A:Cross-references: EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865
C:Genetics:
A:Gene: GDB: CENPE
A:Cross-references: GDB:361164; OMIM:117143
A:Map position: 4q24-q25
C:Superfamily: centromere protein E; kinesin motor domain homology
C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
P:7-335/Domain; kinesin motor domain homology <KNOT>
P:86-93/Region; nucleotide-binding motif A [P-loop]
P:486-2183/Domain; coiled coil #status predicted <COI>
P:92/Binding site: ATP (lys) #status predicted

Alignment Scores:
Pred. No.: 9,35e-09 Length: 2663
Score: 266.50 Matches: 172
Percent Similarity: 38.30% Conservative: 126
Best Local Similarity: 22.11% Mismatches: 305
Query Match: 7.83% Indels: 175
DB: 1 Gaps: 30

US-09-602-362E-15 (1-2030) x S28261 (1-2663)
```

```
QY      12 AAAGATGCTCTCTGAAGGCTAACTGCGGAATGAAAGTTTCTTATCCAACTAAAGCCTTA 71
Db      1111 LysGluGluGluLeuSerArgThrCysAspArgLeuAlaGluGluLysLeuLys 1130
QY      72 GAATGTGATGACATGCAAACTTTCAAAGCAGAGCCTCCGAGAGCCATCTGCCTTCGAG 131
Db      1131 Glu-----LysSerGlnGlnLeuGlnLysGlnGlnLeuLeuAsnValGlnGlu 1148
QY      132 CCTGCCATTGAATGCAAAAG-----TCTGTTCCAAATAAAGCCTTGGATTGAAG 182
Db      1149 GluMetSerGluMetGlnLysLysIleAsnGluIleGluAsnLeuLysAsnGluLys 1168
QY      183 AAT---GAACAAACATTGACAGCAGATGACATCTCCATCAGAATCCAAACAAAG--- 236
Db      1169 AsnLysGluLeuThrLeuGluHisMetGluThrGluArgLeuGluAlaGlnLysLeu 1188
QY      237 -----GACTATGAAGAAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTCAC 290
Db      1189 AsnGluAsnTyrgluGlu-----ValLysSerIle-----ThrLysGluArg 1202
QY      291 AAGGATCTGTGTTTACCACAGGCTACACATCAAAAGAAATAGATAAAATAAATGGAATA 350
Db      1203 LysValLeuLysGlnLeuGlnLysSerPheGluThrGluArgAspHisLeuArgGlyTy 1222
QY      351 TTAGAAGAGTCTCTGTGATAATGATGTTTCTGAAGCTCCCTGCAGAAATGAAAGTTTCT 410
Db      1223 IleArgGlu----- 1225
QY      411 ATTCCAACCTAAAGCCTTAGAATTTGATGGACATGCAAACTTCAAGACAGAGCCTCCGAG 470
Db      1226 -----IleGluAlaThrGlyLeuGlnThrLysGluGluLeuLysIleAlaHisIleHis 1243
QY      471 AAGCCATCTCCCTTCGAGCCTGCCATTGAAATCAAAAGTCTGTTCCAAATAAAGCCTTG 530
Db      1244 LeuLysGluHisGlnGlnThrIleAspGluLeuArgArgSerValSerGluLysThrAla 1263
QY      531 GAATTGAGAAATGAACAAACATTG----- 554
Db      1264 GlnIleLeuAsnThrGlnAspLeuGluLysSerHisThrLysLeuGlnGluLeuPro 1283
QY      555 -----AGAGCAGATCAGATGTTCTCCCTTCAGAATCAAAACAAAGAGGTTGAAGAA 605
Db      1284 ValLeuHisGluGlnGlnGluLeuLeuProAsnValLysLysValSerGluThrGlnGlu 1303
QY      606 AATCTTGGGATTCCTGAGACTCTCCGTGAGACTGTTTTCACAGAGGATGTGTGTATACC 665
Db      1304 ThrMetAsnGluLeuGluLeuLeuGlnSerThrThrLysAspSerThrThrLeu 1323
```

| | | | |
|----|------|---|------|
| QY | 666 | ARGGCTACACATCAAAAAGAAATG-----GATAAAATAAGTGAAAAATTAGAAAGAT | 716 |
| Db | 1324 | AlaArgIleGluMetGluArgLeuArgLeuAsnGlnLysPheGlnGluSerGlnGluGlu | 1343 |
| QY | 717 | TCACTAGCCGTATCAAAAATCTTGATACAGTTCATCTCTGCAAGAGCAAGGGAACCT | 776 |
| Db | 1344 | IleLysSerLeuThrLysGluArgAspAsnLeuLysThrIleLysGluAlaLeuGluVal | 1363 |
| QY | 777 | CAA-----AAAGATCACTGTGAACAACAGTACAGAAAAAATGGAAACAAATGAA | 824 |
| Db | 1364 | LysHisaspGlnLeuLysGluHisIleArgGluThrLeuAlaLysIleGlnGluSerGln | 1383 |
| QY | 825 | AGNAG-----TTTGTGTACTGAAAGAAACTGTCAGAGCAACAAAGAAATAAAA | 875 |
| Db | 1384 | SerLysGlnGluInSerLeuAsnMetLysGlnLysAspAsnGluThrThrLysIleVal | 1403 |
| QY | 876 | TCAGAGTTAGAACACCAAAA-----GTTAAATGGGAACAA | 911 |
| Db | 1404 | SerGluMetGluGlnPheLysProLysAspSerAlaLeuLeuArgIleGluIleGluMet | 1423 |
| QY | 912 | GAGCTCTGCAGTGTGAGATTCATCTTAAACCAAGAGACAGAGAG----- | 956 |
| Db | 1424 | LeuGlyLeuSerLysArgLeuGlnGluSerHisAspGluMetLysSerValAlaLysGlu | 1443 |
| QY | 957 | -----AGAGAAATCCGATATATTAATGAA | 993 |
| Db | 1444 | LysAspAspLeuGlnArgLeuGlnGluValLeuGlnSerGluSerAspGlnLeuLysGlu | 1463 |
| QY | 984 | AAAATTAGGGAAGAATTAGGAAGA----- | 1007 |
| Db | 1464 | AsnIleLysGluIleValAlaLysHisLeuGluThrGluGluLeuLysValAlaHis | 1483 |
| QY | 1008 | -----ATCGNAGACGACAT-----AGGAAA | 1028 |
| Db | 1484 | CysCysLeuLysGluGlnGluGluThrIleAsnGluLeuArgValAsnLeuSerGluLys | 1503 |
| QY | 1029 | GAGTTTAGAA-----GTCAACAACAATCTGAACAG-----GCTCTC | 1064 |
| Db | 1504 | GluThrGluIleSerThrIleGlnLysGlnLeuGluAlaIleAsnAspLysLeuGlnAsn | 1523 |
| QY | 1065 | AGATACAGATATAGAAATTGAAGAGTGTAGAAGTAATTTGATCAGGTTCTCACACT | 1124 |
| Db | 1524 | LysIleGlnGluIleThrGluLysGlnGluGlnLeuAsnIleLysGlnIleSerGluVal | 1543 |
| QY | 1125 | CATGAAATGAAATTTATCTTACATGAAATTCATGTTGAAAAAGGAAATGCCATG | 1184 |
| Db | 1544 | GlnGluAsnValAsn-----GluLeuLysGln | 1552 |
| QY | 1185 | CTAAATCTGGAATAGCCACTGAAACCAATACAGAAAGGAAATAATACTTTT | 1244 |
| Db | 1553 | PheLysGluHisArgLysAlaLysAspSerAlaLeuGlnSerIleGluSerLysMetLeu | 1572 |
| QY | 1245 | GAGGACATTAAAGTTTAAAAAGAAAGAACTCTGAACCTTCAGATGACCCCTAAACTGAAA | 1304 |
| Db | 1573 | GluLeuThrAsnArgLeuGlnGluSerGlnGluGluIleGlnIleMetIleLysGluLys | 1592 |
| QY | 1305 | GAGGAATCATTAACAAAAGGCGCATCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCT | 1364 |
| Db | 1593 | GluGlu-----MetLysArgValGlnGluAlaLeuGlnIleGluArgAspGlnLeuLys | 1610 |
| QY | 1365 | GAGAACACAATG-----CTCACTCTCAATTTGAAGGAAAAACAAGACAAAGAAATACTAGAG | 1421 |
| Db | 1611 | GluAsnThrLysGluIleIleValAlaLysMetLysGluSerGlnGluLysGlu-Tyr----- | 1628 |
| QY | 1422 | GCAGAAATTTGAATCACACCATCTCAGACTGGCTTCTCTGTGTACAAAGACCATGATCAATT | 1481 |
| Db | 1629 | -GlnPheLeuLysMetThrAlaValAsn-----GluThrGlnGluLysMe | 1643 |
| QY | 1482 | GTGACATCAAGAAAAAGTCAGAACCTGCTTCCACATTCAGGAGATGCTTGTTCGAA | 1541 |
| Db | 1643 | tCysGluIleGluHisLeuLysGluGlnPheGluThrGlnLysLeuAsnLeuGluAsnIle | 1663 |
| QY | 1542 | AGAAAAATGAATGTTGATGTGAGTAGTACCCGATATATAACAATGAGGTGCTCCATCAACC | 1601 |

Db 1663 eGlu- ||| |||||-ThrGluAsnIleArgLeuThrGlnIleLeuHisGluAs 1677

Qy 1602 ACTTCTGAAGCTCAAAAGGAATCC-ANAGCCTAAAAATTAA 1643

Db 1677 nLeuGluGluMetArgSerValThrIysGluAArgAspLeuArgSerValGluGluTh 1697

Qy 1644 TCTCAATTATGCAGGAGACTCTTAACAGAGAAAATACATTGGTTTCAGGAACATGCACAAA 1703

Db 1697 rLeuLysValGluArgAspGlnLeuLysGluAsn-----LeuArgGluThrIleThr-A 1715

Qy 1704 GAGAC-----CAACGTGAACAACAG---TGTCAAATGAAGGAGCTGAACACATGT 1751

Db 1715 rgAspLeuGluLysGlnGluLeuLysIleValHisMetHisLeuLysGluHisGlnG 1735

Qy 1752 NTCAAANCAGAACA-----CATNATGTGAACAACACACTGANCAGC 1793

Db 1735 luThrIleAspLysLeuArgGlyIleValSerGluLysThrAsnGluIleSerAsnMetG 1755

Qy 1794 AGGAGTCTCTAGATCAGAAATATTTCACATAACAAGCAAAAATATGTGCTTCAACAGC 1853

Db 1755 lnyAspLeuGluHisSerAsnAspAlaLeuLysAlaGlnAspLeuLysIleGlnGluG 1775

Qy 1854 AATTA-----GTTCTATGCATANGAAGCTGACACAAAAGCAAGATAACAATTGAT 1907

Db 1775 luLeuArgIleAlaHisMetHisLeuLysGlnGlnGluGlnGluThrIleAsp----- 1791

Qy 1908 NTTCAATTCTTGAGAGAAATGCNCATCATCTTCTAAAAGAGAAAATGAGGAGATAT 1967

Db 1792 -----LysLeuArgGlyIleValSerGluLysThrAspLysLeus 1805

Qy 1968 TTNATTACNATACCATTAAAAAACCCGPATATTTCATATCGAAAAAA 2017

Db 1805 exAsnMetGlnLysAspLeuGluAsnSerAsnAlaLysLeuGlnGluLys 1821

RESULT 11

S26710
spindle pole body protein NUF1 - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: protein D9476.3; protein YDR356W
C:Species: Saccharomyces cerevisiae
C>Date: 07-May-1993 #sequence revision 07-May-1993 #text_change 23-Mar-2001
C:Accession: S26710; S34288; A49455; S61152
R:Mirzayan, C.; Copeland, C.S.; Snyder, M.
J. Cell Biol. 116, 1319-1332, 1992
A:Title: The NUF1 gene encodes an essential coiled-coil related protein that is
A:Reference number: S26710; MUID:92176232; PMID:1541631
A:Accession: S26710
A:Molecule type: DNA
A:Residues: 1-944 <MUR>
A:Cross-references: EMBL:Z11582; NID:g4069; PIDN:CAA77668.1; PID:g4070
A>Note: The authors translated the codon GAG for residue 206 as Asp and CTG for
F:Kilmartin, J.V.; Dyos, S.B.; Kershaw, D.; Finch, J.T.
submitted to the EMBL Data Library, June 1993
A:Description: A spacer element in the Saccharomyces cerevisiae spindle pole body
A:Reference number: S34287
A:Accession: S34288
A:Molecule type: DNA
A:Residues: 1-944 <KL>
A:Cross-references: EMBL:X73297; NID:g312173; PID:g312175
F:Kilmartin, J.V.; Dyos, S.B.; Kershaw, D.; Finch, J.T.
J. Cell Biol. 123, 1175-1184, 1993
A:Title: A spacer protein in the Saccharomyces cerevisiae spindle pole body who
A:Reference number: A49455; MUID:94064779; PMID:7503995
A:Accession: A49455
A:Molecule type: DNA
A:Residues: 1-22 <KR2>
A:Cross-references: EMBL:X73297
R:Dü, Z.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of *S. cerevisiae* cosmid 9476.
A:Reference number: S61148
A:Accession: S61152
A:Molecule type: DNA

| | | | |
|----|------|--|------|
| Qy | 924 | GTGAGATTGACTTTTAAACCAAGAAGAGAGAGAAATCCCGATATATTAAATGAA | 983 |
| Db | 497 | -----LysGluArgLupheAsnTyrLysIleSerGluSer | 508 |
| Qy | 984 | AAATATTAGGGAAGNATTAGGAAGNATCGAAGACGACGATAGGAAGAGTGTAGAGTGAAA | 1043 |
| Db | 509 | LysLeuGluAspGluLysThrLeuAsnGluLysIleSerAsnLeuAlaAlaGluAsn | 528 |
| Qy | 1044 | CAACAACTTGAACAGGCTCTCAGAATAACAAGATATAGAAATTGAAGAGTGTAGAAAGTAAT | 1103 |
| Db | 529 | SerGlnLeuLysAsn-----LysIleGluAsp----- | 537 |
| Qy | 1104 | TTGAATCAGGTTTCTCACATCATGAAATAATGAAAAATTATCTTACATGAAAAATTGCGATG | 1163 |
| Db | 538 | ---AsnSerThrAlaThrHisMetLysGluAsnTyr | 549 |
| Qy | 1164 | TTGAAAAGCAAAATGGCATGCTTAAACTGGAATAGCCACACTGAAACACCAATACCAG | 1223 |
| Db | 550 | ---GluLysGlnLeuGluSerLeuArgLysAspIleGluLutyrLysGluSerAlaLys | 568 |
| Qy | 1224 | GAAGAAGAAATAAATCTTTGAG----- | 1247 |
| Db | 569 | AspSerGluAspLysIleGluGluLeuLysIleArgIleAlaGluAsnSerAlaLysVal | 588 |
| Qy | 1248 | -----GACATTAAAGTTTTAAAGAAAAGAAATCTGAATTCAGATTCAGATG | 1289 |
| Db | 589 | SerGluLysArgSerLysAspIleLysGlnLysAspGluGlnIleSerAspLeuThrGln | 608 |
| Qy | 1290 | ACCTAAACTGAAAGAGGAATCAATTAACATAAAGGCGCATCT----- | 1331 |
| Db | 609 | AsnLeuLysLeuGlnGluAspGluIleSerSerLeuLysSerIleIleAspArgTyrLys | 628 |
| Qy | 1332 | -----CAATATAGTGGCGAGCTTAAA | 1352 |
| Db | 629 | LysAspPheAsnGlnLeuLysSerGluGlnSerAsnIleGlnHisAspLeuAsnLeuGln | 648 |
| Qy | 1353 | GTCTCATGCTCAGAACACAAATGCTCACCTTCTAAATTGAAGGAAAAACAAGACAAGAA | 1412 |
| Db | 649 | IleLeuAsnLeuGluAsnLysLeuIleGluSerGluAspGluLeuLysSerLeuArgAsp | 668 |
| Qy | 1413 | ATACTAGGCGAAGATTGAATCACCATCTCTAGACTGGCTTCTGCTGTACAGACCAT | 1472 |
| Db | 669 | SerGlnLysIleGluIleGluAsnTyrLysArgLysTyrAsnAsnLeuSerLeuGluAsn | 688 |
| Qy | 1473 | GATCAAAATTGTGACATCAAGAAAAAGTCAAGAACCTCTTTCACATCTGCAGGAGATGCT | 1532 |
| Db | 689 | AspArgLeuLeuThrGluLysGluSer-----AlaSerAsp--- | 700 |
| Qy | 1533 | TGTTTGCAGAAAAAATGAATGTTGATGTAGTAGTACCGATATATAACAATCAGGTGCT | 1592 |
| Db | 700 | ----- | 700 |
| Qy | 1593 | CCATCAACCACTTCTCAAGCTCAAAGGAAATCCANAGCCTAAAAATTAATCTCAATTA | 1652 |
| Db | 701 | -----LysGluArgLupheAsnTyrLysIleSerGluSer | 708 |
| Qy | 1653 | TGCAGGAGATGCTCTAAGAG-----AAAATACATTGG-----TTTTCAGGAACATGCACAA | 1703 |
| Db | 708 | euAsnArgLysLeuAspGluMetAspLysGluLysTyrAsnLeuGlnSerLysGluL | 728 |
| Qy | 1704 | GAGACCAACGTGAACACAGTGTCAATAGGAAGCACTGAAACACATGTTTCAANCGAAC | 1763 |
| Db | 728 | ysTyrLysIleGluLeuGln-----LysValIleThrAlaAsnAspArgLeuArgGluL | 747 |
| Qy | 1764 | AAAGTATGTGAACAAACACACTGANCAGCAGGAGTCTCTAGATCAGAAATATTATTCAA- | 1822 |
| Db | 747 | ysGluGluLeuAsnGluAsnSerAsnIleArgIleMetGluAspLysMetThrArgI | 767 |
| Qy | 1823 | -----CTACAAAGCAAAAATATGTGCTTCAC | 1850 |
| Db | 767 | leLysLysAsnTyrLeuSerGluIleThrSerLeuGlnGluAsnArgLeuGluGluG | 787 |

Db 926 GlnLeuAlaArgSerIleAlaGluGluGlnThrSerAspLeuGluGluGluSerIleMet 945
Qy 1422 GCAGAAATGAATCAACCATCTAGACTGGCTTCGTGTGAAGACCATCATCAATTT 1481
Db 946 LysGluLeuGlu-----IleLysGluMetMetAlaArgHisLysGlnGlu 960
Qy 1482 GTGACATCAAGAAAAGTCAAGAACCTGCTTTCCACATTCGAGAGATGCTTTGTGAA 1541
Db 961 LeuThrGluLysAspAlaThr-----IleAlaSerLeuGluGluThrAsn 975
Qy 1542 AGAAAAATGAATGTTGATGTGAGTAGTAGTACCGATATATACAAATGAGGTGCTCCATCAACC 1601
Db 976 ArgThrLeuThrSerAspValAlaAsnLeu-AlaAsnGluLysGluGluLeuAsnLys 995
Qy 1602 ACTTCTGAGCTCAAGAAATCCNAAGCTTAAATAATATCAATATGCGAGAGA 1661
Db 995 sLeuLysGluAlaGlnGluGlnLeuSerArgLeuLys---AspGluGluIleSerAlaAl 1014
Qy 1662 TGCTCTAAGA-----GAAAATACATTTGGTTTCAGGAACATGCACAAAGACCAAC 1712
Db 1014 alaIleLysAlaGlnPheGluLysGlnLeuLeu-----ThrGluArgThrLe 1030
Qy 1713 GTGAACAACAGTGTCAAAATGAAGAGCTGAACACATGNTCAAACGAAGAAGATNATG 1772
Db 1030 u-LysThrGlnAlaValAsnLysLeuAlaGluIleMetAsnArgLysGluProValLysA 1050
Qy 1773 TGAACAAACACACT-----GANCAGCAGGAGCTCTAGATCAGAAATTA---TTTCAAC 1823
Db 1050 rGlyAsnAspThrAspValArgLysGluLysGluAsnArgLysLeuHisMetGluL 1070
Qy 1824 TACAAGCAAAAATATGTGGCTTCAACAGCAATTAGTTTCATGCACATAANGAAAGCTGAC 1883
Db 1070 euLysSerGluArgGluLysLeuThrGlnMetIleLys-----T 1084
Qy 1884 ACACAAAGCAAGATACATTCATTTTCATTTCTTGAGAGAAATGCNATCATCTTT- 1942
Db 1084 yrGlnLysGluLeuAsnGluMetGlnAlaGlnIleAlaGluGluSerGlnIleArgIleG 1104
Qy 1943 -----CTAAAGAGAGAAAATGAGAGATATTNATTACNATAACCATTTAA 1988
Db 1104 LuLeuGlnMetThrLeuAspSerLysAspSerAspIleGluGlnLeuArgSerGlnLeuG 1124
Qy 1989 AAAACCCGTATATTCATA 2008
Db 1124 InAlaLeuHisIleGlyLeu 1130
RESULT 13
A41604
Myosin heavy chain, smooth muscle, long splice form - rabbit
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C:Accession: A41604; A33501
R:Babji, P.; Kelly, C.; Periasamy, M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10676-10680, 1991
A:Title: Characterization of a mammalian smooth muscle myosin heavy chain gene: complete
A:Reference number: A41604; PMID:92073350; PMID:1961735
A:Accession: A41604
A:Molecule type: mRNA
A:Residues: 1-1972 <BAB>
A:Cross-references: GB:M77812
R:Nagai, R.; Kuro-O, M.; Babji, P.; Periasamy, M.
J. Biol. Chem. 264, 9734-9737, 1989
A:Title: Identification of two types of smooth muscle myosin heavy chain isoforms by cDN
A:Reference number: A33501; PMID:8925535; PMID:2722872
A:Accession: A33501
A:Molecule type: mRNA
A:Residues: 1455-1972 <NAG>
A:Cross-references: GB:J04833; NID:g165519; PIDN:AAA31407.1; PID:g165520
A:Experimental source: smooth muscle
A>Note: examination by Southern blotting for the regions of difference between this isofo
active splicing
C:Superfamily: myosin heavy chain; myosin motor domain homology

C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
F:188-771/Domain: myosin motor domain homology <XMO>
F:178-185/Region: nucleotide-binding motif A (P-loop)
F:559-572/Region: actin binding #status predicted
F:633-647/Region: actin binding #status predicted
F:844-1938/Domain: coiled coil #status predicted <COI>
F:844-1284/Region: S2
F:1285-1972/Region: light meromyosin
F:1939-1972/Domain: carboxyl-terminal <CBT>
F:123/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
F:184/Binding site: ATP (Lys) #status predicted
F:701,711/Active site: Cys #status predicted
Alignment Scores:
Pred. No.: 1.86e-08 Length: 1972
Score: 261.50 Matches: 128
Percent Similarity: 40.00% Conservative: 118
Best Local Similarity: 20.81% Mismatches: 215
Query Match: 7.68% Indels: 154
DB: 1 Gaps: 18
US-09-602-362E-15 (1-2030) x A41604 (1-1972)
Qy 528 TTGGAATTGAAGAAATGAACAA----- 548
Db 827 LeuLysLeuArgAsnTrpGlnTrpTrpArgLeuPheThrLysValLysProLeuLeuGln 846
Qy 549 ---ACATTGAGACAGATCAGATGTTCCCTTCAGATCAAAACAAAGAGGTGAAGAA 605
Db 847 ValThrArgGlnGluGluMetGlnAlaLysGluAspGluLeuGlnLysLysGlu 866
Qy 606 AATTCTTGGGATCTGAGAGTCTCCGTGAGACTGTTTCACAGAGAGGTGTGTGTACCC 665
Db 867 ArgGlnGlnLysAlaGluSerGluLeuGlnGlnGlnLysHisThrGlnLeu--- 885
Qy 666 AAGCTACACATCAAAAGAAATGGATAAAATAGTGGAAATTAGAAGATTCACATGAC 725
Db 886 -----SerGluGluLysAsnLeuLeuGlnGlnGlnLeuGlnAlaGluThrGlu 901
Qy 726 CTA-----TCAAAAATCTTG 740
Db 902 LeuTrpAlaGluAlaGluGluMetArgValArgLeuAlaAlaLysLysGlnGluLeuGlu 921
Qy 741 GATACAGTTTCATTCTTGTGAAGAGCAAGGGAACCTTCAAAAGAGTCACTGTGTGAACACGT 800
Db 922 GluIleLeuHisGluMetGluAlaArgLeuGluGluGluAspArgGlyGlnGlnLeu 941
Qy 801 ACAGAAAAATGGAACAAATGAAAAAGAGAGTGTGTGTACTGAAAGAACTGTGAGAA 860
Db 942 GlnAlaGluArgLysLysMetAlaGlnGlnMetLeuAspLeuGluGlnLeuGluGlu 961
Qy 861 GCAAAAGAAATAAAATCACAGTTA----- 884
Db 962 GluGluAlaAlaArgGlnLysLeuGlnLeuGluLysValThrAlaGluAlaLysLys 981
Qy 885 -----GAGAACCAAAAGATTAAATGGGAACAA 911
Db 982 LysLeuGluAspAspIleLeuValMetAspAspGlnAsnAsnLysLeuSerLysGluArg 1001
Qy 912 GAGCTCTGAGTGTGAGATTGACT-----TTAAACCAAGAGAGAGAGAG 956
Db 1002 LysLeuLeuGluGluArgIleSerAspLeuThrAsnLeuAlaGluGluGluLys 1021
Qy 957 AGAAGAAATCCGATATATTAAATGAAAAATTTAGGGAAGAAATTAGGAAGATTCGAAGAG 1016
Db 1022 AlaLysAsnLeuThrLysLeuLysAsnLysHisGluSerMetIleSerGluLeuGluVal 1041
Qy 1017 CAGCATAGGAAGAGGTAGAGTGAACACAACTTGAACAGCTCTCAGATACAGAT 1076
Db 1042 ArgLeuLysLysGluGluLysSerArgGlnGlnLeuGluLysLysLysArgLysMetAsp 1061
Qy 1077 ATAGAATTGAAGAGTGTAGAAAGTAATTTCAATCAGGTTTCTCAGCTCATGAAATGAA 1136
|||

```

Db      1062 GlyAlaSerAsp----- 1066
QY      1137 AATTATCTCTACATGAAATTCATGTTGTAAGAAAGAAATTCCTGATGCTAAAGCTGAA 1196
Db      1067 -----LeuHisGlu----- 1077
QY      1197 ATAGCCCACTGAAACACCAATACCAAGGAAAGAAATTAATATCTTTGAGGACATTAAG 1256
Db      1078 IleAlaGluLeuLysMetGlnLeuAlaLysGluGluGluAlaLeuAla 1097
QY      1257 AATTTAAAGAAAGAAATGCTGAATCTCAGATCAGCCCTAAAG----- 1298
Db      1098 ArgLeuGluAspGluThrSerGlnLysAsnAlaLeuLysLysIleArgGluLeuGlu 1117
QY      1299 -----CTGAAAGAGGAATCAATTAATAAGGGCATCTCAATATATAGTGGG 1343
Db      1118 GlyHisIleSerAspLeuGlnGluAspSerGluArgAlaAlaArgAsnLysAla 1137
QY      1344 -----CAGCTTAAGATTCATAGCTGAG----- 1367
Db      1138 GluLysGlnLysArgAspLeuGluGluLeuGluAlaLeuLysThrGluLeuGluAsp 1157
QY      1368 -----AACCAATGCTCACTCTTAATTAAGGAAAGAAACAAAGAAAGAAATA--- 1415
Db      1158 ThrLeuAspThrThrAlaThrGlnGlnGluLeuArgAlaLysArgGluGlnGluValThr 1177
QY      1416 -----CTAGAGGAGAAATGTAATCACACCATCTAGAGCTGCTTCTGCT 1460
Db      1178 ValLeuLysLysAlaLeuAspGluGluThrArgSerHisGluAlaGlnValGlnGluMet 1197
QY      1461 GTACAGACCATGATCAATATGTCATCAAGAAAGAAAGTCAAGAACTGCTTCCACATT 1520
Db      1198 ArgGlnLysHisThrGlnValValGluGluLeuThrGluGlnGluGlnPheLysArg 1217
QY      1521 GCAGGAGATGCTTGTTCGAAAGAAATATCAATGTTGATGTGAGTAGTACCATATATAA 1580
Db      1218 AlaLysAlaAsnLeuAspLysThrLysGlnThrLeuGluLysGluAsnAlaAspLeu-Al 1237
QY      1581 CAATGAG-----GTGCTCCATCAACCATCTTCTGAAGCTCAAGGAAATCCANAAGCT 1634
Db      1237 aGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGluHisLysLysLysLe 1257
QY      1635 AAAAATTAATCTC-----AATTATGAGGAGATGCTCTAAGAGAAATATACATT 1682
Db      1257 uGluValGlnLeuGlnGluLeuGlnSerLysCysSerAspGlyGluArgAlaArgAlaG 1277
QY      1683 GGTTCAGGACATGTCACAAAGACCAACGACGTAAGAACACAGTGT-----CAAATCA 1733
Db      1277 uLeuAsn-AspLysValHisLysLeuGlnAsnGluValGluSerValThrGlyMetLeu 1297
QY      1734 AGGAAGCTGAA-----CACATGT 1751
Db      1297 erGluAlaGluGlyLysAlaLysLeuAlaLysGluValAlaSerLeuGlySerGlnL 1317
QY      1752 NTCAAACGACCAAGATNATGTGAACAAACACACTGACGAGGAGTCTTAGATACAGA 1811
Db      1317 euGlnAspThrGlnGluLeuGlnGlnGluThrArgGlnLysLeuAsnValSerThrL 1337
QY      1812 AATTATTTCAACTACAAAGCAAAATATGTGGCTTCAACAGCAATTAGTTCATGCACATA 1871
Db      1337 ysLeuArgGlnLeuGluAspGluArgAsnSerLeuGlnGlnGluLeu-----A 1353
QY      1872 ANGAAGCTGACAAACAAAGCAAGATPAACAATTGATNTTCATNTCTTCTGAGAGGAATG 1931
Db      1353 spGluGluMetGluAlaLysGln--Asn-LeuGluArgHisIleSerThrLeuAsnIle 1371
QY      1932 CNCATCATCTTCTAAAGGAAAGAAAGGAGATATTTNATACNATAACCATTTAAAAA 1991
Db      1372 GlnLeuSerAspSerLysLysLysLysGlnAspPheAlaSerThrVal----- 1387
QY      1992 ACCCGTATATTTCAATATGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2030
Db      1388 ---GluSerLeuGluGluGlyLysLysArgPheGlnLys 1399

```

```

RESULT 14
B43402
myosin heavy chain-B, neuronal - chicken
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Gallus gallus (chicken)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C:Accession: B43402; A43402
R:Takahashi, M.; Kawamoto, S.; Adelstein, R.S.
J. Biol. Chem. 267, 17864-17871, 1992
A:Title: Evidence for inserted sequences in the head region of nonmuscle myosin specific
yosin.
A:Reference number: A43402; MUID:92388144; PMID:1355479
A:Accession: B43402
A:Molecule type: mRNA
A:Residues: 1-2007 <TA>
A:Cross-references: GB:M93676; NID:G212448; PIDN:AAA48989.1; PID:G212452
A:Note: The sequence of residues 212-221 and 632-652 and the corresponding nucleotide se
A:Accession: A43402
A:Molecule type: mRNA
A:Residues: 1-211;222-631;653-2007 <TA2>
A:Cross-references: GB:M93676; NID:G212448; PIDN:AAA48989.1; PID:G212449
A:Note: sequence extracted from NCBI backbone (NCBIN:1128664)
C:Comment: Alternatively spliced segments 1 and 2 are found exclusively in nonmuscle myos
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
F:1-2007/Product: myosin heavy chain-B, neuronal #status predicted <MYN>
F:1-211;222-631;653-2007/Product: myosin heavy chain-B, nonmuscle #status predicted <MYS>
F:88-802/Domain: myosin motor domain homology <MMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)
F:212-221/Region: alternatively spliced segment 1 #status experimental
F:559-593/Region: actin binding #status predicted
F:632-652/Region: alternatively spliced segment 2 #status experimental
F:692-714/Region: actin binding #status predicted
F:875-2007/Domain: coiled coil #status predicted <COI>
F:875-1315/Region: S2
F:1316-2007/Region: light meromyosin
F:129/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
F:184/Binding site: ATP (Lys) #status predicted
F:732,742/Active site: Cys #status predicted
F:1954/Binding site: phosphate (Thr) (covalent) #status predicted
F:1987/Binding site: phosphate (Ser) (covalent) #status predicted

Alignment Scores:
Pred. No.: 1,99e-08 Length: 2007
Score: 261.00 Matches: 144
Percent Similarity: 38.81% Conservative: 116
Best Local Similarity: 21.49% Mismatches: 238
Query Match: 7.67% Indels: 172
DB: Gaps: 18

US-09-602-362E-15 (1-2030) x B43402 (1-2007)
QY 378 TTTCTGAAGCTCCCTCGACAATGAAAGTTTCTTCAACTAAAGCTTAGAATGATG 437
Db 369 PheThrLysValLysProLeuLeuGlnValThrArgGlnGluGluGlnAlaLys 888
QY 438 GACATGCAAACTTTCAAAGCAGAGCCCTCCGAGAGCCATCTCGCTTCGAGCCGCTG 497
Db 889 AspGluGluLeuMetLysValLys-----GluLysGlnThrLysValGluAlaGluLeu 906
QY 498 GAATGCAAAAGTCTGTTCCAATAAAGCTTCGAATTTGAAGAT-----GAACAA 548
Db 907 GluGluMetGluArgLysHisGlnGlnLeuLeuGluGluLysAsnIleLeuAlaGluGln 926
QY 549 ACATTGAGAGCAGATCAGATGTTCCCTTCA-----GAATCA 584
Db 927 LeuGlnAlaGluThrGluLeuPheAlaGluAlaGluGluMetArgAlaArgLeuAlaAla 946
QY 585 AAACAAAGAGGTTGAAGAAATTTCTTGGATTCTTGAGAGTCTCCGTGAGACTGTTCA 644
Db 947 LysLysGlnGluLeuGluGluGluLeuLeuHisAspLeuGluSerArgValGluGluGlu 966

```

645 CAGAAGGATGTGTGTACCCCAAGGCTACACATCAAAAGAAATCGATAAAATAGTGA 704
Db : : : : :
967 GluArgAsnGlnIleLeu-----GlnAsnGluLysLysMetClnGly 981
QY : : : : :
705 AAATTAGAGATTCAACTAGCTATCAAAATCTTGGATACAGTTCATCTTGTGAAGA 764
Db : : : : :
982 HisIleGlnAsp-----LeuGluGlnGlnLeuAspGluGluGluGluValAlaArgGln 998
QY : : : : :
765 GCAAGGGAACTTCAAAAGATCACTGTGAACAACCTACAGGAATAATCGAACAAATGAA 824
Db : : : : :
999 LysLeuGlnLeuGluLysValThrAlaGluAlaLysIleLysLysMetClnGlu----- 1016
QY : : : : :
825 AAGAAGTTTGTGTACTGAAAGAACTGTGAGAACCAAGAAATATAATACAGCTTA 884
Db : : : : :
1017 -----GluIleLeuLeuGluAspGlnAsnSerLysPhe 1028
QY : : : : :
885 GAGAACCAAAAGTTAAATGGGAAACAAGAGCTCTCGAGTGTGAGATTGACTTTAAACCA 944
Db : : : : :
1029 LeuLysGluLysLeuMetClnAspArgIleAlaGluCysThrSerGlnLeuAlaGlu 1048
QY : : : : :
945 GAAGAGAGAGAGAGAAATCCGATATATTAATGAAATAATTAGGAAAGAAATTAGGA 1004
Db : : : : :
1049 GluGluGluLysAlaLysAsnLeuAlaLysLeuLysAsnLysGlnGluMetMetIleThr 1068
QY : : : : :
1005 AGAATCGAAGACAGCAGTGAAGAGTAGTAGTGAACACAACTTGAACAGCTCTC 1064
Db : : : : :
1069 AspLeuGluGluArgLeuLysGluGluLysThrArgGlnGluLeuGluLysAlaLys 1088
QY : : : : :
1065 AGAATACAAGATATAGATTGAAGAGTGTAGAAAGTAAATTTGAATCAGCTTCTCACACT 1124
Db : : : : :
1089 ArgLysLeuAspGlyGluThrThrAspLeuGlnAspGlnIleAlaGlu----- 1104
QY : : : : :
1125 CATGAAATGAAATATCTCTTACATGAAATTCGATGTTGAAAGAAAGAAATGCCANG 1184
Db : : : : :
1105 -----LeuGlnAlaGlnIleGluGlu 1111
QY : : : : :
1185 CTAAGCTGGAATAGCCACACTGAACACCAATACCCAGAA----- 1226
Db : : : : :
1112 LeuLysIleGlnLeuAlaLysLysGluGluGluGlnAlaLeuAlaLeuArgGlyAsp 1131
QY : : : : :
1227 -----AAGGAAATTAATTAATCTTGGAGCATTAAGATTTTAAAGAAAGAAAT 1274
Db : : : : :
1132 GluGluAlaValGlnLysAsnAlaLeuLysValIleArgGluLeuGlnAlaGlnIle 1151
QY : : : : :
1275 GCTGACTTCAGATCACCTTAACCTGAAGAGGAATCATTAACATAAGGCGATCTCAA 1334
Db : : : : :
1152 AlaGluLeuGlnGluAspLeuGluSerGluLysAlaSerArgAsnLysAlaGluLysGln 1171
QY : : : : :
1335 -----TATAGTGGGAGCTTAAAGTTCTGATAGCTGAG----- 1367
Db : : : : :
1172 LysArgAspLeuSerGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeuAsp 1191
QY : : : : :
1368 AACACATGCTCCTCTTAATTTGAAGGAAACAAAGACAAAGAAATACTA----- 1418
Db : : : : :
1192 ThrThrAlaAlaGlnGlnGluArgThrLysArgGluGlnGluValAlaGluLeuLys 1211
QY : : : : :
1419 -----GAGGAGAGAAATTTGAATCATCACCATCTCTAGA 1448
Db : : : : :
1212 LysAlaIleGluGluThrLysAsnHisGluAlaGlnIleGlnGluIleArgGlnArg 1231
QY : : : : :
1449 CTGGCTCTGCTGTCAAGACATCATCAAAATGTGACATCAA----- 1491
Db : : : : :
1232 HisAlaThrAlaLeuGluGluLeuSerGluGlnLeuGluGlnAlaLysArgPheLysAl 1251
QY : : : : :
1492 -----GAAAGCTCAAGACCTCTCTTCACATTCGAGGAGATGCTGTTTCAAGA 1544
Db : : : : :
1251 asnLeuGluLysAsnLysGlnGlyLeuGluSerAspAsnLysGluLeuAlaCysGluVal 1271
QY : : : : :
1545 AAAATGAATGTTGATGTGAGTAGTACCGATATATAACAATGAGTGTCTCCATCAACCACT 1604
Db : : : : :
1271 lLys-----ValLeuGlnGlnVally 1278
QY : : : : :
1605 TTCTGAAGCTCAAGGAAATCCNAAGCCTA----- 1635

Db : : : : :
1278 sAlaGluSerGluHisLysArgLysLysLysLeuAspAlaGlnValGlnGluLeuThrAlaLys 1298
QY : : : : :
1635 ----- 1635
Db : : : : :
1298 sValThrGluGlyGluArgValGluLeuAlaGluLysAlaAsnLysLeuGlnAs 1318
QY : : : : :
1636 -----AAATTAATCTCAATTATGC 1655
Db : : : : :
1318 nGluLeuAspAsnValSerSerLeuLeuGluGluAlaGluLysGlyLysPheAl 1338
QY : : : : :
1656 AGGAGATGCTCTTAAGAAATATACATTGGTTTCAGGAACATGCACAAAGACCAACGTG 1715
Db : : : : :
1338 aLysAspAlaAlaSerLeuGluSerGlnLeuGlnAspThrGlnGluLeuGlnGluG 1358
QY : : : : :
1716 AAACACAGTCTCAATCAAGAGAGCTCAACACATGTCATCAACGACCAAGATNATGTGA 1775
Db : : : : :
1358 lThrArgGlnLysLeuAsnLeuSerSerArgIleArgGlnLeuGluGluGluLys--A 1377
QY : : : : :
1776 ACAACACACTGANCACGAGAGTCTCTAGATCAAGAAATTTATTTCACAAAGCAAAA 1835
Db : : : : :
1377 sAsnLeuGlnGlnGlnGlnGluGluGluGluAlaArgLysAsnLeuGluLysGlnM 1397
QY : : : : :
1836 ATATGCTGCTTCAACAGCAATTAATTCATGCACATCAANGAAAGCTGAC----- 1883
Db : : : : :
1397 etLeuAlaLeuGlnAlaGlnLeuAlaGluAlaLys-LysLysValAspAspLeuGly 1416
QY : : : : :
1884 -----AAACAAAGCAAGATAACATTCAT-----NTT 1910
Db : : : : :
1417 ThrIleGluGlyLeuGluGluAsnLysLysLeuLysAspMetGluSerLeuSer 1436
QY : : : : :
1911 CATNTCTTGAGAGGAAATGCNCATCATCTCTTAAAGAGAGAAAA----- 1956
Db : : : : :
1437 GlnArgLeuGluGlu-LysAlaMetAlaTyAspLysLeuGluLysThrLysAsnArgLe 1456
QY : : : : :
1957 -----TGAGGAGATTTTATTACNATAACCATTTTAAAAACCCGTATATTCAA 2006
Db : : : : :
1456 uGlnGlnGluLeuAspAspLeuMetValAspLeuAspHisGlnArgGlnIleValSerAs 1476
QY : : : : :
2007 TATGAGAAAAAANAANAANA 2028
Db : : : : :
1476 nLeuGluLysGlnLysLys 1483
RESULT 15
JC5421
smooth muscle myosin heavy chain 2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 02-Feb-2001
C:Accession: JC5421
R:Hasegawa, K.; Arakawa, E.; Oda, S.; Matsuda, Y.
Biochem. Biophys. Res. Commun. 232, 313-316, 1997
A:Title: Molecular cloning and expression of murine smooth muscle myosin heavy chains.
A:Reference number: JC5420; MUID:97242182; PMID:9125171
A:Accession: JC5421
A:Molecule type: mRNA
A:Residues: 1-1938 <HAS>
A:Cross-references: DDBJ:D85924; NID:G1945079; PIDN:BAAL19691.1; PID:G1945080
A:Experimental source: smooth muscle
C:Comment: This protein plays a role in smooth muscle cell contraction.
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: nucleotide binding; P-loop
F:88-771/Domain: myosin motor domain homology <MMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)
Alignment Scores:
Pred. No.: 3,01e-08 Length: 1938
Score: 258.00 Matches: 132
Percent Similarity: 39.33% Conservative: 113
Best Local Similarity: 21.19% Mismatches: 233
Query Match: 7.58% Indels: 145
DB: 2 Gaps: 17
US-09-602-362E-15 (1-2030) x JC5421 (1-1938)

QY 417 ACTAAAGCCCTTAGAATTGTATGGACATGCAAACTTTCAAAGCAGACGCTCCCGAGAGCCA 476
Db 839 ThrLysValLysProLeuLeuGlnValThrArgGlnGluGluMetGlnAlaLysGlu 858
QY 477 TCTGCTTCGAGCCTGCCATTGCAATGCAAAAGTCTGTTCCAAATAAGACCTTGAATTG 536
Db 859 GluGluMetGlnLysIleThrGluArgGlnGlnLysAlaGluThrGluLeuLysGluLeu 878
QY 537 AAGAATGAAACAACATTTGAGCAGCAGATCAGATGTTCCCTTCAGATCAAAACAAAGAG 596
Db 879 GluGlnLysHisThrGlnLeuAlaLysLysGlnGluLysThrLeuLeuGlnGlnLeuGlnAla 898
QY 597 GTTGAAGAAATCTCTGGGATCTGAGAGTCTCCGTGAGACTGTTTCCAGAGAGATGTG 656
Db 899 GluThrGluLeuTyraAlaGluSerGluGluMetArg----- 910
QY 657 TGTGTACCAAGCCTACACATCAAAAGAAATGGATAAAATAGTGGAAATTTAGAAGAT 716
Db 911 ---ValArgLeuAlaLysLysGlnGluLeuGluGluLe----- 923
QY 717 TCAACTAGCCTATCAAAATCTTTGATACAGTTCATTTCTGTGAAGAGCAAGGAACTT 776
Db 924 -----LeuHisGluMetGluAlaArgLeuGluGlu 933
QY 777 CAAAAGATCACTGTGAACAACGCTACAGGAAATAATGGACAAATGAAAGAAAGTGTGT 836
Db 934 GluGluAspArgGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGlnMetLeu 953
QY 837 GTACTGAAAAGAACTGTGCAAGCAAAAGAAATAAATCACTTA----- 884
Db 954 AspLeuGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeuGluLys 973
QY 885 -----GAG 887
Db 974 ValThrAlaGluAlaLysIleLysLysLeuGluAspAspIleLeuValMetAspGln 993
QY 888 AACCAAAAAGTTAAATGGGAACAAGAGCTCTGCAGTGTGAGATTGACT----- 935
Db 994 AsnSerLysLeuSerLysGluArgLysLeuLeuGluGluArgValSerAspLeuThrThr 1013
QY 936 ---TTAAACCAAGCAAGAGAGAGCAAGAAATCGCGATATATTAATGAAATAATTAGG 992
Db 1014 AsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysSerLysHisGlu 1033
QY 993 GAAGAATTAGGAAGATCCGAAGCAGCAGCATAGGAAGAGTTAGAAAGTGAACAACAACCTT 1052
Db 1034 SerMetIleSerGluLeuGluValArgLeuLysLysGluLysSerArgGlnGluLeu 1053
QY 1053 GAACAGGCTCTCAGAATACAGATATAGAAATGAGAGGTGAGAAAGTAATTGTAATCAG 1112
Db 1054 GluLysLeuLysArgLysLeuGlu-----GlyAsp 1063
QY 1113 GTTTCACACTCATCAAAATGAAATTTATCTTTACATGAAATTTGCATGTGAAAGAG 1172
Db 1064 AlaSerAspPheHisGlu----- 1069
QY 1173 GAAATTCCTCAATGTTAAATGGAAATGACCACTGAAACACACATACACAGGAAGAGAA 1232
Db 1070 GlnIleAlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGlu 1089
QY 1233 AATAAATCTTTCAGGACATTAAGATTTTAAAGAAAGAAAGTCTGAACTTCAGATGACC 1292
Db 1090 GluGluLeuGlnAlaAlaLeuAlaArgLeuAspGluGluIleAlaGlnLysAsnAla 1109
QY 1293 CTAAAA-----CTGAAAGAGGAATCATTAAC 1319
Db 1110 LeuLysLysIleArgGluLeuGluGluHisIleSerAspLeuGlnGluAspLeuAspSer 1129
QY 1320 AAAGGGCATCTCAATATAGTGG-----CAGCTTAA 1352
Db 1130 GluArgAlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGluGluLeuGlu 1149

QY 1353 GTTCTGATAGCTGAG-----AACACAATGCTCACTTCTTAATTGAAG 1394
Db 1150 AlaLeuLysThrGluLeuGluAspThrLeuAspSerIleAlaThrGlnGlnGluLeuArg 1169
QY 1395 GAAAAACAACACAAAGAAATA-----CTAGAGCAGAAATTTGAATCA 1436
Db 1170 AlaLysArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSer 1189
QY 1437 CACATCTTAGACTGCTCTCTGTGTACAGACCATGTCATTAATTTGACATCAAGAAAA 1496
Db 1190 HisGluAlaGlnValGlnGlnMetArgGlnLysHisThrGlnAlaValGluGluLeuThr 1209
QY 1497 AGTCAAGAACTCTTTCCACATTTGAGGAGATGCTTTGTCAAAGAAAAATGAATGT 1556
Db 1210 GluGlnLeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysSerLysGlnThrLeu 1229
QY 1557 GATGTAGTAGTAGTACCATATATAACAATGAG-----GTGCTCCATCAACCTTTCTCA 1610
Db 1230 GluLysGluAsnAlaAspLeu-AlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnG 1249
QY 1611 AGCTCAAGGAAATCCANAAGCCTTAAATAATTAATCTC-----AATTATGCAGG 1658
Db 1249 uValGluHisLysLysLysLysLeuGluValGlnLeuGlnAspLeuGlnSerLysCysSe 1269
QY 1659 AGATGCTCTTAAGAGAAATACATTTGCTTTCAGGAACATGCCAAAGAGACCAACGTGAA 1718
Db 1269 rAspGlyGluArgAlaArgAlaGluLeuSer-AspLysValHisLysLeuGlnAsnGluV 1289
QY 1719 CACAGTGT-----CAAATGAAGAAAGCTGAA----- 1744
Db 1289 alGluSerValThrGlyMetLeuAsnGluAlaGluGlyLysAlaIleLysLeuAlaLysA 1309
QY 1745 -----CACATGTTNTCAAANCGAACACAGATNATGTGAACAAACACACTG 1787
Db 1309 spValAlaSerLeuGlySerGlnLeuGlnAspThrGlnLeuLeuGlnGluGluThra 1329
QY 1788 ANCAGCAGGAGTCTCTAGATCAGAAATTTTCAACTACAAAGCAAAATATGTGCTTC 1847
Db 1329 rgGlnLysLeuAsnValSerThrLysLeuArgGlnLeuGluAspGluArgAsnSerLeuG 1349
QY 1848 AACAGCAATTAGTTTCATGCACATAANGAAAGCTGCACAACAAAGCAAGATAACAATTGAT 1907
Db 1349 InAspGlnLeu-----AspGluGluMetGluAlaLysGln---Asn-LeuGlu 1363
QY 1908 NTTCAATNTCTTGAGAGGAAATGCGCATCATCTTCTTAAAGAGAAATATGAGGAGATAT 1967
Db 1364 ArgHisValSerThrLeuAsnIleGlnLeuSerAspSerLysLysLysLeuGlnAspPhe 1383
QY 1968 TTNATTACNATAACCATTTTAAAAAACCCGTATATTTCAATATATCGAAAAAANAANA 2027
Db 1384 AlaSerThrIle-----GluValMetGluGluGlyLysArgLeuGln 1398
QY 2028 AAA 2030
Db 1399 Lys 1399

Search completed: July 15, 2004, 09:47:57
Job time : 105.671 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 08:50:37 ; Search time 95.0076 Seconds
(without alignments)

13356.780 Million cell updates/sec

Title: US-09-602-362E-15

Perfect score: 3403

Sequence: 1 ctctgtgcgttaagatggtt.....gaaataaaanaaaaaaa 2030

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1285345 seqs, 312560633 residues

Total number of hits satisfying chosen parameters: 2570690

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/p/US09602362/runat_15072004_093627_22071/app.query.fasta_1.10325
-DB=Published Applications AA -OFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNIT5=bits -SPART=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pcr -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09602362 @CGN_1_1_719 @runat_15072004_093627_22071
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| ----- | | | | | |

RESULT 1

US-10-058-270A-4

; Sequence 4, Application US/10058270A

; Publication No. US20040029114A1

; GENERAL INFORMATION:

; APPLICANT: Mack, David H.

; APPLICANT: Gish, Kurt C.

; APPLICANT: Afar, Daniel

; APPLICANT: Eos Biotechnology, Inc.

; TITLE OF INVENTION: Methods of Diagnosis of Breast Cancer, Compositions and

; FILE REFERENCE: 018501-005210US

; CURRENT APPLICATION NUMBER: US/10/058,270A

; CURRENT FILING DATE: 2002-01-24

; PRIOR FILING DATE: 2001-01-24

; PRIOR FILING DATE: 2001-01-24

; PRIOR FILING DATE: 2001-02-02

; PRIOR FILING DATE: 2001-02-02

; PRIOR FILING DATE: 2001-04-09

; PRIOR FILING DATE: 2001-04-09

; PRIOR FILING DATE: 2001-04-09

; PRIOR FILING DATE: 2001-04-09

; PRIOR FILING DATE: 2001-04-09

; PRIOR FILING DATE: 2001-05-04

; PRIOR FILING DATE: 2001-05-04

ALIGNMENTS

| 1 | 3272 | 96.2 | 1341 | 12 | US-10-058-270A-4 |
|----|--------|------|------|----|--------------------|
| 2 | 3272 | 96.2 | 1341 | 13 | US-10-007-805-565 |
| 3 | 3272 | 96.2 | 1341 | 14 | US-10-076-622-565 |
| 4 | 3272 | 96.2 | 1341 | 14 | US-10-177-293-334 |
| 5 | 3272 | 96.2 | 1341 | 14 | US-10-124-805-565 |
| 6 | 3272 | 96.2 | 1349 | 13 | US-10-007-805-573 |
| 7 | 3272 | 96.2 | 1349 | 14 | US-10-076-622-573 |
| 8 | 3272 | 96.2 | 1349 | 14 | US-10-124-805-573 |
| 9 | 3261 | 95.8 | 1002 | 9 | US-09-604-287A-475 |
| 10 | 3261 | 95.8 | 1002 | 10 | US-09-551-621-475 |
| 11 | 3261 | 95.8 | 1002 | 13 | US-10-007-805-475 |
| 12 | 3261 | 95.8 | 1002 | 14 | US-10-076-622-475 |
| 13 | 3261 | 95.8 | 1002 | 14 | US-10-124-805-475 |
| 14 | 3261 | 95.8 | 1095 | 13 | US-10-007-805-493 |
| 15 | 3261 | 95.8 | 1095 | 14 | US-10-076-622-493 |
| 16 | 3261 | 95.8 | 1095 | 14 | US-10-124-805-493 |
| 17 | 3255 | 95.7 | 1013 | 13 | US-10-007-805-553 |
| 18 | 3255 | 95.7 | 1013 | 14 | US-10-076-622-553 |
| 19 | 3255 | 95.7 | 1013 | 14 | US-10-124-805-553 |
| 20 | 2599 | 76.4 | 512 | 16 | US-10-181-663-16 |
| 21 | 2256.5 | 66.3 | 1239 | 13 | US-10-007-805-577 |
| 22 | 2256.5 | 66.3 | 1239 | 14 | US-10-076-622-577 |
| 23 | 2256.5 | 66.3 | 1239 | 14 | US-10-124-805-577 |
| 24 | 2184 | 64.2 | 445 | 9 | US-09-825-301-29 |
| 25 | 2184 | 64.2 | 445 | 9 | US-09-604-287A-473 |
| 26 | 2184 | 64.2 | 445 | 10 | US-09-551-621-473 |
| 27 | 2184 | 64.2 | 445 | 13 | US-10-007-805-473 |
| 28 | 2184 | 64.2 | 445 | 14 | US-10-076-622-473 |
| 29 | 2184 | 64.2 | 445 | 14 | US-10-124-805-473 |
| 30 | 2184 | 64.2 | 445 | 14 | US-10-033-527-29 |
| 31 | 2017 | 59.3 | 432 | 9 | US-09-825-301-18 |
| 32 | 2017 | 59.3 | 432 | 9 | US-09-604-287A-181 |
| 33 | 2017 | 59.3 | 432 | 9 | US-09-339-338-181 |
| 34 | 2017 | 59.3 | 432 | 10 | US-09-551-621-181 |
| 35 | 2017 | 59.3 | 432 | 13 | US-10-007-805-181 |
| 36 | 2017 | 59.3 | 432 | 14 | US-10-076-622-181 |
| 37 | 2017 | 59.3 | 432 | 14 | US-10-124-805-181 |
| 38 | 2017 | 59.3 | 432 | 14 | US-10-033-527-181 |
| 39 | 2017 | 59.3 | 432 | 15 | US-10-441-893-181 |
| 40 | 1785 | 52.5 | 1225 | 14 | US-10-177-293-332 |
| 41 | 1597.5 | 46.9 | 650 | 9 | US-09-825-301-25 |
| 42 | 1597.5 | 46.9 | 650 | 9 | US-09-604-287A-469 |
| 43 | 1597.5 | 46.9 | 650 | 10 | US-09-551-621-469 |
| 44 | 1597.5 | 46.9 | 650 | 13 | US-10-007-805-469 |
| 45 | 1597.5 | 46.9 | 650 | 14 | US-10-076-622-469 |

Sequence 4, Appli
Sequence 565, App
Sequence 565, App
Sequence 334, App
Sequence 565, App
Sequence 573, App
Sequence 573, App
Sequence 573, App
Sequence 475, App
Sequence 475, App
Sequence 475, App
Sequence 475, App
Sequence 493, App
Sequence 493, App
Sequence 493, App
Sequence 493, App
Sequence 553, App
Sequence 553, App
Sequence 16, Appl
Sequence 577, App
Sequence 577, App
Sequence 29, Appl
Sequence 18, Appl
Sequence 181, App
Sequence 181, App
Sequence 181, App
Sequence 181, App
Sequence 181, App
Sequence 181, App
Sequence 332, App
Sequence 25, Appl
Sequence 469, App
Sequence 469, App
Sequence 469, App

; PRIOR APPLICATION NUMBER: US 60/294,443
 ; PRIOR FILING DATE: 2001-05-29
 ; NUMBER OF SEQ ID NOS: 141
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 4
 ; LENGTH: 1341
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-058-270A-4

Alignment Scores:

Pred. No.: 4,45E-233 Length: 1341
 Score: 3272.00 Matches: 656
 Percent Similarity: 97.49% Conservative: 3
 Best Local Similarity: 97.04% Mismatches: 13
 Query Match: 96.15% Indels: 5
 DB: 12 Gaps: 0

US-09-602-362E-15 (1-2030) x US-10-058-270A-4 (1-1341)

QY 9 GTTAAAGATGCTTTCTGAGGCTAACTCGGAAATCAAGATTTCTATTCCAACTAAAGCC 68
 Db 666 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 685
 QY 69 TTAGAATTGATGACATGCAAACTTTCAAAGCAGAGCCCTCCCGAGAAGCCATCTGCCTTC 128
 Db 686 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 705
 QY 129 GAGCCTGCCATGAAATGCAAAAGTCTGTTCCAAATAAAGCCCTTGAATTTGAAGATGAA 188
 Db 706 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 725
 QY 189 CAAACATTGAGACGACATGAGATACCTCCCATCAGAAATCCAAACAAAGAGCTATCAAGAA 248
 Db 726 GlnThrLeuArgAlaAspGluLeuLeuProSerGluSerLysGlnLysAspTyrGluGlu 745
 QY 249 AGTCTTGGATTTCTGAGAGTCTCTGTGAGACTGTTTCACAGAAGATGTGTGTTTACC 308
 Db 746 SerSerTyrAspSerGluSerLeuLysGluThrValSerGlnLysAspValCysLeuPro 765
 QY 309 AAGGCTACACATCAAAAGAAATAGATAAAATAAATGGAATTTAGAAGAGTCTCTGAT 368
 Db 766 LysAlaThrHisGlnLysGluLeuAspLysIleAsnGlyLysLeuGluGluSerProAsp 785
 QY 369 AATGATGGTTTCTGAAGCTCCCTGCAGATCAAAAGTTTCTATTCCAACTAAAGCCTTA 428
 Db 786 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 805
 QY 429 GAATTCATGACATCAAACTTTCAAGCAGAGCCCTCCCGAGAAGCCATCTGCCTTCGAG 488
 Db 806 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 825
 QY 489 CTTGCCATTGAATGCAAAAGTCTGTCCAAATAAAGCCCTTGAATTTGAAGATGAACAA 548
 Db 826 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 845
 QY 549 ACATTGAGACGACATCAGATGTTCCTTCAGATCAAAACAAAGAGTTTGAAGAAAT 608
 Db 846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsn 865
 QY 609 TCTTGGATTCTCAGAGTCTCCGTGAGACTGTTTCACAGAAGGATGTGTGTATCCCAAG 668
 Db 866 SerTyrAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
 QY 669 GCTACACATCAAAAGAAATGGAATAAATAGTGAAGAAATTAAGATTTCAACTAGCCTA 728
 Db 886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 905
 QY 729 TCAAAATCTTGATACAGTTCTTCTGTGAAGACCAAGGAACTTCAAAAGATCAC 788
 Db 906 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 925
 QY 789 TGTGAACACGTTACAGAAATAATGGAAACAAATGAAAGAGTTTTGTGTACTGAAAGAG 848

Db 926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 945
 QY 849 AAATGTCAGAGCAAAAGAAATAAATACAGTTAGAGAACCAAAAAGTTAAATCGGAA 908
 Db 946 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTyrGlu 965
 QY 909 CAAGAGCTCTGCAGTGTGAGATTGACTTTAAACCAAGAAAGAGAGAGAGAAATGCC 968
 Db 966 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 985
 QY 969 GATATATTAAATCAAAAATTTAGCGAAGATTAGGAAGAAATCGAAGCAGCATAGAAA 1028
 Db 986 AspIleLeuAsnGluLysIleArgGluLeuGlyArgIleGluGlnGlnHisArgLys 1005
 QY 1029 GAGTTAGAAGTGAACCAACACTTGAACAGGCTCTCAGAAATACAAGATATAGAATTGAAG 1088
 Db 1006 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 1025
 QY 1089 AGTGTAGAAAATTAATTTGAATCAAGTCTTCACACTCATGAAATGAAAATTTATCTCTTA 1148
 Db 1026 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 1045
 QY 1149 CATGAAAATTTGCATGTTGAAAAGGAATTCCTCATCTTAAACTGAAATAGCCACATG 1208
 Db 1046 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 1065
 QY 1209 AAMCACCATAACACAGAAAGAAATAAATACCTTTGAGGACATTAAGATTTTAAAGAA 1268
 Db 1066 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 1085
 QY 1269 AAGAATCTGAACTTCAGATGACCTTAAACTGAAAGAGAGATCATTAATAAGGGCA 1328
 Db 1086 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 1105
 QY 1329 TCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGACCAATGCTCACTTCTAAA 1388
 Db 1106 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1125
 QY 1389 TTGAAGAAAACCAAGCAAAAGAAATACTAGAGCGCAGAAATGAATCACAACCATCTAGA 1448
 Db 1126 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 1145
 QY 1449 CTGGCTTCTCTGTACAGACCATGATCAATTTGTCACATCAAGAAAAGTCAAGAACCT 1508
 Db 1146 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1165
 QY 1509 GCTTTCCACATTCGAGGAGATGCTTGTTCGAAAGAAAATGAATCTTGATGTGAGTAGT 1568
 Db 1166 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1185
 QY 1569 ACCGATATATAACANTGAGGTCTCCATCAACACCTTCTGAAGCTCAAGGAAATCCAN 1628
 Db 1186 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 1205
 QY 1629 AAGCTAAAAATTAATCTCAATTTATCGAGGAGATGCTCTAAGAGAAAATACATGTTTTC 1688
 Db 1205 SerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSer 1225
 QY 1689 AGGAACATGCACAAAGAGACCAACGTTGAACACAGTGTCAATGAAGAGAGCTGAACACA 1748
 Db 1225 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHis 1245
 QY 1749 TGTNTCAAAACGCAAGATNATGTGTAACAAACACACTGANCAGCAGGAGTCTCTAGATC 1808
 Db 1245 eTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1265
 QY 1809 AGAAATTTTCAACTCAAAAGCAAAAATATGTGGCTTCAACAGCAATTAAGTTATGTCAC 1868
 Db 1265 LysLeuPheGlnLeuGlnSerLysAsnMetTyrLeuGlnGlnGlnLeuValHisAlaHis 1285
 QY 1869 ATAANGAAAGCTGACAAACAAAGCAAGATAACAATTTGATNTTTCATTTCTTGAGAGAAA 1928

Db 1285 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1304
QY 1929 ATGCN-CATCATCTTCTAAAGAGAAAATGAGGAGATATTNATACNATAACCATTTA 1987
Db 1305 MetGlnHisLeuLeuLysGluLysAsnGluIlePheAsnTyrAsnAsnHisLeu 1324
QY 1988 AAAAACCCTGATATTCATATGGAAGAAAAAANAANA 2027
Db 1325 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1337
RESULT 2
US-10-007-805-565
; Sequence 565, Application US/10007805
; Publication No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 565
; LENGTH: 1341
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-007-805-565
Alignment Scores:
Pred. No.: 4,456-233 Length: 1341
Score: 3272.00 Matches: 656
Percent Similarity: 97.4% Conservative: 3
Best Local Similarity: 97.04% Mismatches: 13
Query Match: 96.15% Indels: 5
DB: 13 Gaps: 0
US-09-602-362E-15 (1-2030) x US-10-007-805-565 (1-1341)
QY 9 GTTAAAGATGTCCTTCTGAAGCTAACTGCGGAATGAAAGTTTCTATTCCAACTAAAGCC 68
Db 666 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 685
QY 69 TTAGAATTGATGACATGCAAACTTCAAGCAGAGCTCCCGAAGCCATCTGCTTC 128
Db 686 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPhe 705
QY 129 GAGCTGCAATGAATGCAAGTCTGTTCCAAATGAAGCTTGGATTGGAATGAA 188
Db 706 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 725
QY 189 CAACATTCAGACGACATGAGATATCTCCCATCAGAAATCAACAAAGAGCTATGAGAA 248
Db 726 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 745
QY 249 AGTTCTGGATCTGAGAGTCTCTGAGACTGTTTCACAGAGGATGTGTGTACCC 308
Db 746 SerSerTrpAspSerGluSerLeuLysGluThrValSerGlnLysAspValCysLeuPro 765
QY 309 AAGGCTACACATCAAAAAGAAATAGATAAAATGAAATGAAATGAAAGAGTCTCTGAT 368
Db 766 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 785

QY 369 AATGATGTTTCTGAAGGCTCCCTGCAGAAATGAAAGTTTCTATTCCAACTAAAGCCTTA 428
Db 786 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 805
QY 429 GAATTGATGACATGCAAACTTCAAGCAGAGCTCCCGAAGCCATCTGCTTCGAG 488
Db 806 GluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu 825
QY 489 CTTGCCATTCAAAATGCAAAAGTCTGTTCAAAATGAAAGCTTGGAAATGAAGAATGAACA 548
Db 826 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 845
QY 549 ACATTGAGACGACATCAGATGTTCCCTTCAGAAATCAAAAAGAGGTTTGAAGAAAT 608
Db 846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLysValGluGluAsn 865
QY 609 TCTTGGATTCGAGAGTCTCGTGAGACTGTTTCACAGAGGATGTGTGTACCCAAG 668
Db 866 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
QY 669 GCTACACATCAAAAAGAAATGGAATAAAGTGGAAATTAGAAGATTCAACTAGCCTA 728
Db 886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 905
QY 729 TCAAAATCTTGGATCAGTTCAATCTTGTGAAAGCAAGGAACTTCAAAAAGATCAC 788
Db 906 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 925
QY 789 TGTGAACACGTCACAGAAAATGGAACAAATGAAAGAAAGTTTGTGTACTGAAAGAA 848
Db 926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 945
QY 849 AAAGTGTGAGAACAAAGAAATAAATCACTTAGAGAACCAAAAGTAAATGGA 908
Db 946 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 965
QY 909 CAAGAGCTCTGAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAGAGATGCC 968
Db 966 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 985
QY 969 GATATATTAATGAAAGAAATTAGGAGAAATTAGGAGAAATCGAAGACGACATAGAAA 1028
Db 986 AspIleLeuAsnGluLysIleArgGluLeuGlyArgIleGluGlnHisArgLys 1005
QY 1029 GAGTTAGAAGTGAACAACTTGAACAGGCTCTCAGATACAGATATAGAAATGAAG 1088
Db 1006 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 1025
QY 1089 AGTGTAGAAAGTAAATTTGAATCAGGTTTCTCACACTCATGAAATGAAATTTATCTCTTA 1148
Db 1026 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 1045
QY 1149 CATGAAATTCATGTTGAAAAAGGAAATTCGATCTTAACTGAAATAGCACACTG 1208
Db 1046 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 1065
QY 1209 AAACACCAATACAGAGAAAGGAAATAAATCTTTGAGACATTAAGATTATTAAGAA 1268
Db 1066 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysLeuLysGlu 1085
QY 1269 AAGAATGCTGAATTCAGATGACCCCTAAACCTGAAAGAGAAATTCATTAATAAGGCA 1328
Db 1086 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 1105
QY 1329 TCTCAATATAGTGGGAGCTTAAAGTTCTGATAGCTGAGAACACATGCTCACTTCTAA 1388
Db 1106 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1125
QY 1389 TTGAAGAAAAACAAGACAAAGAAATCTAGAGCGCAAAATTCGAATCACACCATCTCTAG 1448
Db 1126 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 1145
QY 1449 CTGGCTTCTCTGTACAGACCATGATCAAAATTTGTGACATCAAGAAAAAGTCAAGAACCT 1508

```

1146 LeuAlaSerAlaValGlnAspHisGlnIleValThrSerArgLysSerGlnGluPro 1165
1509 GCTTTCCACATTCAGAGATGCTGTTTGCAGAAAGAAATCAATGTTGATGTAGTAGT 1568
1166 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1185
1569 ACCGATATATACATGAGGTCTCCATCAACACACTTTCTGAAGCTCAAGGAATCCAN 1628
1186 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 1205
1629 AAGCCTAAATAATTAATCAATATCCAGAGATGCTCTAAGAGAGAAATACATTTGTTTC 1688
1205 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSer 1225
1689 AGGAACATGCACAAAGAGACCAACGTAACACAGTGTCAATGAAGAGAGCTGAACACA 1748
1225 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 1245
1749 TGNNTCAAAACGAAACAGATNATGTGAACAAACACACTGANCAGCAGGAGTCTCTAGATC 1808
1245 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1265
1809 AGAAATTAATTCACACTACAAAGCAAAATATGTGCTTCAACAGCAATTAGTTCATGCAC 1868
1265 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 1285
1869 ATAAGAAACCTGACACAAAGCAAGATTAACAAATGATTTCAATTTCTTCTGAGAGGAAA 1928
1285 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1304
1929 ATCCN-CATCATCTTCTAAAGAGAGAAATGAGGAGATATTTNATACNATAACCATTTA 1987
1305 MetGlnHisHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1324
1988 AAAAAACCGTATTTCAATATGGAAGAAAAAANAANA 2027
1325 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1337

RESULT 3
US-10-076-622-565
; Sequence 585, Application US/10076622
; Publication No. US20030023036A1
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C11
; CURRENT APPLICATION NUMBER: US/10/076,622
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 565
; LENGTH: 1341
; TYPE: PRI
; ORGANISM: Homo sapiens
US-10-076-622-565

Alignment Scores:
Pred. No.: 4,45e-233 Length: 1341
Score: 3272.00 Matches: 656
Percent Similarity: 97.49% Conservative: 3
Best Local Similarity: 97.04% Mismatches: 13
Query Match: 96.15% Indels: 5
DB: 14 Gaps: 0

US-09-602-362E-15 (1-2030) x US-10-076-622-565 (1-1341)
9 GTTAAAGATGGTCTTCTGAGGCTACTCGGAATCAAGATTTCTATCCCACTAAAGCC 68
666 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 685

```

```

69 TTAGAATTGATGACATGCAAACTTTTCAAGCAGAGCCCTCCGAGAGCCATCTGCTTC 128
686 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGlnLysProSerAlaPhe 705
129 GAGCCTGCCATTCAAAATGCTGTTTCCAAATAAAGCCCTTGGAAATGAAGATGAA 188
706 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 725
189 CAACATTGAGACAGATGAGATCTCCCATCAGATCCAAATCCAAACAAAGAGACTATGAGAA 248
726 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 745
249 AGTCTTGGGATCTGAGAGTCTCTGTGAGACTGTTTTCACAGAGGATGTGTGTTTACC 308
746 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 765
309 AAGCTACACATCAAAAGAAATAGATATAAATAATGGAATAATTAGAGAGTCTCTGAT 368
766 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluSerProAsp 785
369 AATGATGCTTTCTGAGGCTCCCTGCAGATGAAAGTTTCTATTCCAACTAAAGCCTTA 428
786 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 805
429 GAATTGATGACATGCAAACTTTTCAAGCAGAGCCCTCCGAGAGCCATCTGCTTCGAG 488
806 GluLeuMetAspMetGlnThrPheLysAlaGluProGlnLysProSerAlaPheGlu 825
489 CCTGCCATTCAAAATGCAAAAGTCTGTTTCCAAATAAAGCCCTTGGAAATGAAGATGAACA 548
826 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysGlnGln 845
549 ACATTGAGACAGATCAGATGTTCCCTTCCAGATCAAAACAAAGAGGTTGAGAGAAAT 608
846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLysValGluGluAsn 865
609 TCTTGGATTTCTGAGAGTCTCCGTGAGACTGTTTTCACAGAGGATGTGTGTACCCAG 668
866 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
669 GCTACACATCAAAAGAAATGCAAAATAAAGTGGAAATTTAGAGATTCACATGACCTTA 728
886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 905
729 TCAAAATCTTGATACAGTTCATTTCTGTAAGAGAGAGGAACTTCAAAAGATCAC 788
906 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 925
789 TGTGAACACGTACAGGAAAAATGGAACAAATCAAAAGAAAGTGTGTGTACTGAAAAAG 848
926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 945
849 AAATCTGTCAAGACAAAGAAATATAATCACAGTTAGAGACCAAAAGTTAAATCGGAA 908
946 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 965
909 CAGAGCTCTGAGTGTGAGATTTGACTTTTAAACCAAGAGAGAGAGAGAAATGCC 968
966 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 985
969 GATATATTAATGAAAAATTTAGCGAAGATTTAGGAAGATTCGAGAGAGAGATGAGAA 1028
986 AspIleLeuAsnGluLysIleArgGluLeuGlyArgIleGluGlnHisArgLys 1005
1029 GAGTTAGAGTGAACACAACTTGAACAGGCTCTCAGATACAGATATAGATTAAG 1088
1006 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 1025
1089 AGTGTAGAAAGTAAATTTGAATTCAGGTTTCTCACACTCATGAAATGAAATTAATCTCTTA 1148
1026 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 1045

```

| | | | |
|----|------|---|------|
| QY | 1149 | CATGAAATATGCTGTTGAAAGAAATATTGCCATGCTAAACATGCGAAATAGCCACACTG | 1208 |
| Db | 1046 | HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu | 1065 |
| QY | 1209 | AAACACCAATACACAGGAAAGGAAATAATACTTTTCAGGACACTTAAGATTTTAAAGAA | 1268 |
| Db | 1066 | LyHisGlnThrGlnGluLysGluAsnLysTyrrPheGluAspIleLysIleLeuLysGlu | 1085 |
| QY | 1269 | AAGAATGCTCAACTTCAGATGACCCCTAAAACCTGAAAGAGGAATCATTAACATAAAGGGCA | 1328 |
| Db | 1086 | LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla | 1105 |
| QY | 1329 | TCTCAATATAGTGGCGACCTTAAAGTCTTGATAGCTGAGAACCAATGCCTCACTTCTTAAA | 1388 |
| Db | 1106 | SerGlnTyrrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys | 1125 |
| QY | 1389 | TTGAAGGAAACAAAGACAAAGAAATACTAGAGGCAGAAATTCGAATCACACCATCCTAGA | 1448 |
| Db | 1126 | LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisProArg | 1145 |
| QY | 1449 | CTGGCTTCCTGCTGTACAAGACCATGATCAAAATGTGTACATCAAGAAAAGCTCAAGACCT | 1508 |
| Db | 1146 | LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro | 1165 |
| QY | 1509 | GCCTTTCACACATTCAGGAGAGATGCTGTGTTTGCAAAGAAAAATGAATGTGATGTGAGTAGT | 1568 |
| Db | 1166 | AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer | 1185 |
| QY | 1569 | ACCGATATATAACATGAGTGTCTCCATCAACCACTTTTCTGAAGCTCAAGGAAATCCAN | 1628 |
| Db | 1186 | Thr-IleTyrrAsnAsnGlnValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLy | 1205 |
| QY | 1629 | AAGCCTAAAAATTAATCTCAATTATGCAGAGAGATGCTCTAAGAGAAATACATGTGGTTTC | 1688 |
| Db | 1205 | sSerLeuLysIleAsnLeuAsnTyrrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe | 1225 |
| QY | 1689 | AGGAACATGCACAAAGAGACCAACGTCGAAACACAGCTGTCAAATCAAGGAAGCTGAAACACA | 1748 |
| Db | 1225 | r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM | 1245 |
| QY | 1749 | TGINTTCAANGCAACAGATNATGTGAACAAACACACTGANCAGCAGGAGTCTCTAGATC | 1808 |
| Db | 1245 | etTyrrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG | 1265 |
| QY | 1809 | AGAAATATTATTTCAACTCAAGCAAAAATATGTGGCTTCAACAGCAATTAGTTCATCCAC | 1868 |
| Db | 1265 | lnLysLeuPheGlnLeuGlnSerLysAsnMetIrrPLeuGlnGlnLeuValHisAlaH | 1285 |
| QY | 1869 | ATAANGAAAGCTGACACAAACAAAGACAGATAACAAATGATNCTCATNTCTTGAGAGAAA | 1928 |
| Db | 1285 | is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys | 1304 |
| QY | 1929 | ATGCN-CATCATCTTCTTAAAGAGAAAAATGAGGAGATATTTNATTTACNATAACCATTTA | 1987 |
| Db | 1305 | MetGlnHisIleLeuLeuLysGluLysAsnGluGluIlePheAsnTyrrAsnAsnHisLeu | 1324 |
| QY | 1988 | AAAAACCCGTATATTTCAATATGAGAAAAAANAANAAAAA | 2027 |
| Db | 1325 | LyAsn-ArgIleTyrrGlnTyrrGluLysGluLysAlaGlu | 1337 |

RESULT 4

US0001
US-021-177-293-334
Sequence 334, Application US/10177293
Publication No. US20030124128A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Glatt, Karen
APPLICANT: Zhao, Xumei
APPLICANT: Gannavarpu, Manjula
APPLICANT: Kamathkar, Shubhang
APPLICANT: Mathens, Maureen
APPLICANT: Myer, Vic
APPLICANT: Wang, Youzhen

QY 429 GAATTGATGACATGCAAACTTCAAGAGAGAGCTCCGAGAGCCATCTGCCTTCGAG 488
Db 806 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 825
QY 489 CTGCGCAATGAAATGCAAAAGCTGTTCCAAATAAGCCTTCGAATTAAGGAATCAACAA 548
Db 826 ProAlaIleGluMetGlnLysSerValProAlaLysAlaLeuGluLeuLysAsnGluGln 845
QY 549 ACATTGAGAGCAGATCAGATGTTCCCTTCAGAAATCAAAACAAAGAGGTTCAAGAAAT 608
Db 846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsn 865
QY 609 TCTGGGATTCGAGAGTCCGTGAGACTGTTTCACAGAGGATGCTGTGTCACCAAG 668
Db 866 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
QY 669 GCTACACATCAAAAGAAATGGATAAAATAAGTGGAAATTAAGAAGATTCAACTAGCCTA 728
Db 886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 905
QY 729 TCAAAATCTTGATACAGTTCACTTCTGTAAGAGAGGAGGAACTTCAAAAGATCAC 788
Db 906 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 925
QY 789 TGTGAACAACGTACAGCAAAATGGAACAAATGCAAAAGAACTTTGTGTACTGAAAGAA 848
Db 926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 945
QY 849 AAATCTCAGAACCAAGAAATAAATCACAGTTAGAGAACCAAAAGATTAAATGGGAA 908
Db 946 LysLeuSerGluAlaLysGluLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 965
QY 909 CAGAGCTCTGCAGTGTGAGTACTTAAACCAAGAGAGAGAGAGAGAAATGCC 968
Db 966 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 985
QY 969 GATATATTAATGAAATAATAGGGAAGAAATAGGAAGAAATCGAAGCAGCATAGGAAA 1028
Db 986 AspIleLeuAsnGlnLysIleArgGluLeuGluArgIleGluGlnHisArgLys 1005
QY 1029 GAGTTAGAGTGAACCAACACTGTAACAGGCTCTCAGATACAGATACAGATTAAGTGAAG 1088
Db 1006 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLysLys 1025
QY 1089 AGTGTAGAAGTAATTTGAATCAGGTTTCTCAGCTCATGAAATGAAATATCTCTTA 1148
Db 1026 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTrpLeuLeu 1045
QY 1149 CATGAAATTCATGTTGAAAGAAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 1208
Db 1046 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 1065
QY 1209 AAACACCAATACCAAGAAAGGAAATAAATCTTTGAGGACATTAAGATTTTAAAGAA 1268
Db 1066 LysHisGlnTrpGlnGlnLysGluAsnLysTrpPheGluAspIleLysIleLeuLysGlu 1085
QY 1269 AAGAATCTGAACTCAGATGACCCCTAAACCTGAAAGAGAAATCAATTAACAAAGGCA 1328
Db 1086 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluSerLeuThrLysArgAla 1105
QY 1329 TCTCAATATAGTGGGAGCTTAAGTCTGATAGCTGAGACACAACTGCTCACTTCTAA 1388
Db 1106 SerGlnTrpSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1125
QY 1389 TTGAAGCAAAACCAAGCAAAATACCTAGAGGCGAGAAATTAATCAACCATCTCTAGA 1448
Db 1126 LeuLysGlnLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 1145
QY 1449 CTGCTCTCTGCTGACAGACCATGATCAATGTCACATCAAGAAAGAGTCAAGAACT 1508
Db 1146 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1165
QY 1509 GCTTTCCACATTGAGGAGATGCTTTGTTGCAAGAAATAATGAATGTTGATGTGAGTAGT 1568

Db 1166 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1185
QY 1569 ACCGATATATAACAATGAGGTGCTCCATCAACACACTTTCTGAAGCTCAAGGAAATCCAN 1628
Db 1186 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 1205
QY 1529 AGCCTAAAATTAATCTCAATTCATCAGAGATGCTCTAAGAGAAATACATTTGGTTTC 1688
Db 1205 sSerLeuLysIleAsnLeuAsnTrpAlaGlyAspAlaLeuArgLysGlnHisLeuValSer 1225
QY 1689 AGCAACATGCACAAAAGAGACCAACAGTCAACAGTGTCAAAATGAAGAGAGCTGAACACA 1748
Db 1225 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 1245
QY 1749 TGNVTCMAANGCAAGATNATGTGAACAAACACTGANCAGCAGGAGTCTCTAGATC 1808
Db 1245 eTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1265
QY 1809 AGAAATTTATTTCAACTACAAAGCAAAATATGTGGCTTCAACAGCAATTAGTTTCATGCAC 1868
Db 1265 lnyLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 1285
QY 1869 ATAANGAAAGCTGACAAACAAAGCAAGATTAACAAATGATNTTCATTTCTTGAGAGAAA 1928
Db 1285 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1304
QY 1929 ATGCN-CATCATCTCTTAAAGAGAAATAGAGAGATATTTNATTACNATAACCACTTTA 1987
Db 1305 MetGlnHisHisLeuLeuLysGluLysAsnGluGluIlePheAsnTrpAsnHisLeu 1324
QY 1988 AAAAACCCGTATATTTCAATATGCAATATGCAAAATAAAATAAAATAAA 2027
Db 1325 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1337
RESULT 5
US-10-124-805-565
; Sequence 565, Application US/10124805
; Publication No. US20030166022A1
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.470C12
; CURRENT APPLICATION NUMBER: US/10/124,805
; CURRENT FILING DATE: 2002-04-15
; NUMBER OF SEQ. ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 565
; LENGTH: 1341
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-124-805-565
Alignment Scores:
Pred. No.: 4,45e-233 Length: 1341
Score: 3272.00 Matches: 656
Percent Similarity: 97.49% Conservative: 3
Best Local Similarity: 97.04% Mismatches: 13
Query Match: 96.15% Indels: 5
DB: 14 Gaps: 0
US-09-602-362E-15 (1-2030) x US-10-124-805-565 (1-1341)
QY 9 GTTAAAGATGCTCTTCTGAAGGCTAACTGCGGAATGAAGCTTCTATTCCAACTAAGCC 68
Db 666 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 685
QY 69 TTACAATTTGATGACATGCAAACTTTTCAAGCAGAGAGCTCCCGAAGAGCCATCTGCTTC 128
Db 686 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 705

QY 129 GAGCTGCGCATGAAATGCAAAAGTCTGTTCCAAATATAAGCCCTTGAAATTGAAGATGAA 188
 Db 706 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 725
 QY 189 CAAACATTCAGAGCAGATGAGATATCTCCCATCAGAAATCCAAACAAAGAGACTATGAAGAA 248
 Db 726 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 745
 QY 249 AGTCTTCGGATTCTCAGAGTCTCTGTGAGACTGTTTCACAGAAGAGTGTGTGTACCC 308
 Db 746 SerSerTrpAspSerGluSerLysGlnThrValSerGlnLysAspValCysLeuPro 765
 QY 309 AGGCTACACATCAAAAGAAATAGATAAAATATAAGTAAATGGAAGAGTCTCCATGAT 368
 Db 766 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 785
 QY 369 AATGATGTTTCTGAGGCTCCCTCGAGAAATGAAAGTTCATTCCAACTAAAGCCCTTA 428
 Db 786 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 805
 QY 429 GAATTGATGACATGCAAACTTTCAAAGCAGAGCCCTCCCGAGAAGCCCATCTGCCCTTCGAG 488
 Db 806 GluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu 825
 QY 489 CCTGCTAATGAATGCAAAAGTCTGTTCCAAATATAAGCCCTTGGAATTCAGAAATGAA 548
 Db 826 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 845
 QY 549 ACATTGAGAGCAGATCAGATGTTCCCTTCAGAAATCAAAACAAAGAGGTTGAAGAAAT 608
 Db 846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsn 865
 QY 609 TCTTGGGATTCAGAGTCTCGTGAGACTGTTTCACAGAGAGTGTGTGTACCCCAAG 668
 Db 866 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
 QY 669 GCTACACATCAAAAGAAATGATATAATAGTGAAGAAATAGAGATTCACATAGCCTA 728
 Db 886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluLysSerThrSerLeu 905
 QY 729 TCAAAATCTTCGATACAGTCTCTGTTGAAGAGCAAGGAACTTCAAAAGATCAC 788
 Db 906 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 925
 QY 789 TGTGAACACGTACAGGAAATGCAAAATGCAAAATGCAAAATGCAAAATGCAAAATGCAAA 848
 Db 926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 945
 QY 849 AAACCTGTGAGAGCAAAAGAAATAGGAAGAAATAGGAAGAAATAGGAAGCAATAGGA 1028
 Db 946 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 965
 QY 909 CAAGAGCTCTGAGTGTGAGATGCTTTAAACCAAGAGAGAGAGAGAGAGAGAGAGAG 968
 Db 966 GlnGluLeuLysSerValArgLeuThrLeuAsnGlnGluGluLysArgAsnAla 985
 QY 969 CATATATTAAATGCAAAATAGGAAGAAATAGGAAGAAATAGGAAGCAATAGGA 1028
 Db 986 AspIleLeuAsnGlnLysIleArgGluLeuGluArgIleGluGluGlnHisArgLys 1005
 QY 1029 GAGTTAGAAGTGAACAAACACTGCAAGGCTCTCAGAAATACAGATATAGAAATGGA 1088
 Db 1006 GluLeuGluValLysGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 1025
 QY 1089 AGTGTAGAAGTAAATTCGAATCAGGTTCTCACACTCATGAATGAAATTAATCTCTTA 1148
 Db 1026 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnLysLeuLeu 1045
 QY 1149 CATGAAATTCGATGTTGAAAGAAATTCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1208
 Db 1046 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 1065

QY 1209 AAACACCAATACACAGGAAAGGAAATATAATCTTTGAGGACATTAAGATTTTAAAGAA 1268
 Db 1066 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 1085
 QY 1269 AAGAACTCTGAATTCAGATGACCCCTAAACTGAAAGAGGAAATCATTAATAAAGGCA 1328
 Db 1086 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluLeuSerLeuThrLysArgAla 1105
 QY 1329 TCTCAATATAGTGGCAGCTTAAAGTCTCATAGCTGAGAACCAATGCTCACTTCTTAA 1388
 Db 1106 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1125
 QY 1389 TTGAAGAAACAAAGACAAAGAAATATACTAGAGCGAGAAATGAATCACACCATCTCTAGA 1448
 Db 1126 LeuLysGlnLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisProArg 1145
 QY 1449 CTGCTTCTCTCTGACAGAGCCATGATCAATTTGTGACATCAAGAAAGCTCAAGAACCT 1508
 Db 1146 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1165
 QY 1509 GCTTTCACATTCGAGGAGATGCTTGTTCGAAAGGAAATGAATGTTGATGTCAGTAGT 1568
 Db 1166 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1185
 QY 1569 ACCGATATATAACATGAGGTGCTCCATCAACACATTCGAGGCTCAAGGAAATCCAN 1628
 Db 1186 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLy 1205
 QY 1629 AAGCTAAAAATTAATCTCAATATGAGGAGATGCTTAAGAGAAATACATTTGTTTC 1688
 Db 1205 SserLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 1225
 QY 1689 AGGACATGCAAAAGAGACCAACCTGAAACACACAGTGTCAAAATGAAGGAGCTGAACACA 1748
 Db 1225 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 1245
 QY 1749 TGTNTCAAAACGACAGATNATGTGAACAAACACACTGACAGCAGGAGCTCTAGATC 1808
 Db 1245 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1265
 QY 1809 AGAAATTTATTCAACTACAAAGCAAAATATGTGGCTTCAACAGCAATTAGTTTCATGCAC 1868
 Db 1265 LysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnLeuValHisAlaH 1285
 QY 1869 ATAANGAAGCTCACACAAACAGATACAAATGATTTTTCATTTCTTCTTGAGAGAA 1328
 Db 1285 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1304
 QY 1929 ATGCN-CATCATCTTCTTAAAGAGAAATGAGGAGATATTTTATTCNATACCATTTTA 1987
 Db 1305 MetGlnHisHisLeuLeuLysGlnLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1324
 QY 1988 AAAACCCGTATATTCATATATGGAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAA 2027
 Db 1325 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1337

RESULT 6

US-10-007-805-573
 ; Sequence 573, Application US/10007805
 ; Publication No. US20020150581A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Yuqiu
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Hepler, William T.
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: Farger, Gary R.
 ; APPLICANT: Vedwick, Thomas S.
 ; APPLICANT: McNeill, Patricia D.
 ; APPLICANT: Durham, Margarita
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
 ; FILE REFERENCE: 210121.470C10
 ; CURRENT APPLICATION NUMBER: US/10/007,805
 ; CURRENT FILING DATE: 2001-12-07

; NUMBER OF SEQ ID NOS: 593
 ; SOFTWARE: fastseq for windows version 4.0
 ; SEQ ID NO 573
 ; LENGTH: 1349
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-007-805-573

Alignment Scores:
 Pred. No.: 4,45e-233 Length: 1349
 Score: 3272.00 Matches: 656
 Percent Similarity: 97.49% Conservative: 3
 Best Local Similarity: 97.04% Mismatches: 13
 Query Match: 96.15% Indels: 5
 DB: 13 Gaps: 0

US-09-602-362E-15 (1-2030) x US-10-007-805-573 (1-1349)

QY 9 GTTAAAGATGCTCTCTCGAAGCTAACTCGGAATGAAAGTTTCTATTCCAACTAAAGCC 68
 Db 674 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 693
 QY 69 TTGAATTTGATGACATGCAAACTTTCAAGCAGAGCCCTCCGAGAGCCATCTCGCTTC 128
 Db 694 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 713
 QY 129 GAGCCTGCCATTCAAAATGCAAAAGTCTGTTCACAAATAAAGCCTTGCAATTGAAGAAATCAA 188
 Db 714 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 733
 QY 189 CAAACATTGAGACGACATGAGATCTCCATCAGAAATCCAAACAAAGACATATGAGNA 248
 Db 734 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspThrGluGlu 753
 QY 249 AGTCTTGGGATCTCAGAGTCTCTGTGAGACTGTTTCACAGAGGATGTGTGTTTACCC 308
 Db 754 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 773
 QY 309 AAGCTACACATCAAAAGAAATAGATAAAATAAATGGAATAATTAGAGAGTCTCTGTAT 368
 Db 774 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluLysProAsp 793
 QY 369 AATGATGGTCTCTGAGGCTCCCTCAGAAATGAAAGTTTCTATTCCAACTAAAGCCTTA 428
 Db 794 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 813
 QY 429 GAATTGATGACATGCAAACTTTCAAAGCAGAGCCTCCGAGAGCCATCTCGCTTCGAG 488
 Db 814 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 833
 QY 489 CCTGCCATTGAATGCAAAAGTCTGTTCACAAATAAAGCCTTGAAATTGAAGAAATGAACA 548
 Db 834 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 853
 QY 549 ACATTGAGACGACATCAGATGTTTCCCTTCAGAAATCAAAACAAAGAGGTTGCAAGAAAT 608
 Db 854 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLysValGluGluAsn 873
 QY 609 TCTTGGGATCTCAGAGTCTCCGTGAGACTGTTTCACAGAGGATGTGTGTCTACCCAG 668
 Db 874 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 893
 QY 669 GCTACACATCAAAAGAAATGGAATAAATGAGTGGAAATTTAGAAATTTCAACTAGCCTTA 728
 Db 894 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 913
 QY 729 TCAAAATCTTGATACAGTTCATTCTGTGAAGAGCAGGGAACCTTCAAAAGATCAC 788
 Db 914 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 933

QY 789 TGTGAACACGTTACAGAAAAATGGAACAAATGAAAGAAAGTTTGTGTACTGAAAAAG 848
 Db 934 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 953
 QY 849 AAATCTGTCAAGAACAAAAAATAAATCACAGTTAGAGAACCAAAAAAGCTTAAATGGGAA 908
 Db 954 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 973
 QY 909 CAAGAGCTCTGAGTGTGAGATTGACTTTAAACCAAGAAAGAGAGAGAGAAATGCC 968
 Db 974 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 993
 QY 969 GATATATTAATCAAAAAATTAGGGAAGAAATTAGGAAGAAATCAAGACAGCATAGAAA 1028
 Db 994 AspileuAsnGlnLysIleArgGluLeuGlyArgIleGluGluGlnHisArgLys 1013
 QY 1029 GAGTTAGACTGAAACAACTTGAACAGGCTCTCAGAAATACAGATATAGAAATTTGAAG 1088
 Db 1014 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspileGluLeuLys 1033
 QY 1089 AGTGTAGAAAGTAATTTGAATCAGGTTCTCACATCATGAAATGAAATTTATCTCTTA 1148
 Db 1034 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTrpLeuLeu 1053
 QY 1149 CATGAAATTTGCATGTTGAAAAAGGAAATTTGCCATCTCTAAACTGAAATAGCACACTG 1208
 Db 1054 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 1073
 QY 1209 AAACCAATACACAGAAAGGAAATTAATCTTTGAGGACATTAAGATTTTAAAGAA 1268
 Db 1074 LysHisGlnTrpGlnGluLysGluAsnLysTrpPheLysAspileLysLeuLysGlu 1093
 QY 1269 AAGAAATCTCAACTTCAGATGACCTTAAACTGAAAGAGGAATCATTAATAAAGGCA 1328
 Db 1094 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluSerLeuThrLysArgAla 1113
 QY 1329 TCTCAATATAGTGGGAGCTTAAAGTTCTGATAGCTGAGACACATGCTCACTCTTAA 1388
 Db 1114 SerGlnTrpSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1133
 QY 1389 TTGAAGCAAAAAACAAGCAAAAGAAATACTAGAGGCAAAATTCGAATCACACCATCTCAGA 1448
 Db 1134 LeuLysGlnLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 1153
 QY 1449 CTGCTTCTCTGTACAGACCATGATCAAAATTTGTGACATCAAGAAAGTCAAGAACCT 1508
 Db 1154 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1173
 QY 1509 GCTTTCACATTCAGGAGATGCTTGTTCGAAAGAAAAATGAATGTGTGAGTAGT 1568
 Db 1174 AlaPheHisIleAlaGlyAspAlaCysLeuGluArgLysMetAsnValAspValSerSer 1193
 QY 1569 ACCGATATATAACATGAGTGTCTCCATCAACACACTTCTGAAGCTCAAGGAAATCCAN 1628
 Db 1194 Thr-IleTrpAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 1213
 QY 1629 AAGCTTAAAAATTAATCTCAATATGAGGAGATGCTCTAAGAGAAAAATACATTGGTTTC 1688
 Db 1213 sSerLeuLysIleAsnLeuAsnTrpAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 1233
 QY 1689 AGGAACATGCAAAAGAGACCAACGTCGAAACACACAGTGTCAAAATGAAGAAAGTGAACACA 1748
 Db 1233 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 1253
 QY 1749 TGTNTCAAAACCAACAGATNATGTGAACAAACACACTGANCAGCAGGAGTCTCTAGATC 1808
 Db 1253 etTrpGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGlnGlnSerLeuAspG 1273
 QY 1809 AGAAATATTTTCAACTACAAAGCAAAAAATATGTGGCTTCAACAGCAATTTAGTTCATGCAC 1868
 Db 1273 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 1293

Db 1074 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 1093
QY 1269 AAGAATCTCAACTTCAGATGACCTAAACTGAAGAGAGAAATCAATAAGGCA 1328
Db 1094 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 1113
QY 1329 TCTCAATATAGTGGGAGCTTAAAGTCTTGATAGCTGAGAACAAATGCTCACTCTAAA 1388
Db 1114 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrWetLeuThrSerLys 1133
QY 1389 TTCAAGGAAACAAAGCAAGAAATATAGAGGCGAGAAATGAATCACACCATCTAGA 1448
Db 1134 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 1153
QY 1449 CTGGCTCTCTGCTCAAGACCATGATCAAAATTTGTCACATCAAGAAAGTCAAGAACT 1508
Db 1154 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1173
QY 1509 GCTTTCCACATTCAGAGAGATGCTTTGTCAGAGAAATGAATGATGTGATGATG 1568
Db 1174 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1193
QY 1569 ACCGATATATAAATGAGTCTCCATCAACACCTTTCTGAAGCTCAAGAAAGTCAAGAACT 1628
Db 1194 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 1213
QY 1629 AAGCTTAAATTAATCTCAATATTCAGAGAGATGCTTAAGAGAAATACATTTGTTTC 1688
Db 1213 SserLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSer 1233
QY 1689 AGGACATGACAAAGAGACCAAGTGAACACAGTGTCAATGAAAGAGAGTGAACACA 1748
Db 1233 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 1253
QY 1749 TGTTTCAAAACGAAACAGATNATGAAACAAACACACTGACAGAGAGTCTAGATC 1808
Db 1253 eTyr-GlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1273
QY 1809 AGAATATTTCACTCAACAAACAAATATGCTTCAACAGCAATTAGTTCATGAC 1868
Db 1273 InLysLeuPheGlnLeuGlnSerLysAsnMetTyrLeuGlnGlnLeuValHisAlaH 1293
QY 1869 ATAANGAAGCTGACAAACAAAGCAAGATACAAATGTATNTTCATNTCTTGAGAGAAA 1928
Db 1293 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1312
QY 1929 ATGCN-CATCATCTTCTAAAGAGAGAAATGAGAGATATTNATTACNATAAACCATTTA 1987
Db 1313 MetGlnHisHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1332
QY 1988 AAAACCCGTATTTCAATATGGAAGAAAAAANAANA 2027
Db 1333 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1345

RESULT 9

US-09-604-287A-475
; Sequence 475, Application US/09604287A
; Patent No. US20020064872A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C7
; CURRENT APPLICATION NUMBER: US/09/604,287A
; CURRENT FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 475

; LENGTH: 1002
; TYPE: PRT
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)-(1002)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-604-287A-475

Alignment Scores:
Pred. No.: 2,78e-232 Length: 1002
Score: 3261.00 Matches: 654
Percent Similarity: 97.19% Conservative: 3
Best Local Similarity: 96.75% Mismatches: 15
Query Match: 95.83% Indels: 5
DB: 9 Gaps: 0

US-09-602-362E-15 (1-2030) x US-09-604-287A-475 (1-1002)

QY 9 GTTAAAGATGCTCTCTGAAGGCTAACTGCGGATGAAGTTCTTATTCCAACTAAAGCC 68
Db 327 ValllysaspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 346
QY 69 TTGAATTTGATGACATGCAAACTTTCAAAGCAGAGCTCCCGAGAGCCATCTGCCTTC 128
Db 347 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPhe 366
QY 129 GAGCTGCCATTGAAATGCAAAAGTCTGTTCCAAATTAAGCCCTTGGAAATGAAGTAA 188
Db 367 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 386
QY 189 CAAACATTTGAGAGCAGATGAGATCTCCCATCAGATCCAAACAAAGAGGACTATGAAGAA 248
Db 387 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 406
QY 249 AGTCTTTGGGATTTCTGAGATCTCTGTGAGACTGTTTCACAGAGGATGTGTGTTACCC 308
Db 407 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 426
QY 309 AAGCTACACATCAAAAGAAATAGATAAATAAATGAAATAGAAATAGAGAGTCTCTGAT 368
Db 427 LysAla**HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluLysProAsp 446
QY 369 AATGATGTTTCTGAAGGCTCCCTGCAGATGAAAGTTTCTTATTCCTCAACTAAAGCTTTA 428
Db 447 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 466
QY 429 GAATTTGATGACATGCAAACTTTCAAAGCAGAGCTCCCGAGAGCCATCTGCCTTCGAG 488
Db 467 GluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu 486
QY 489 CCTGCCATTGAAATGCAAAAGTCTGTTCCAAATTAAGGCTTGGAAATGAAGATGAACAA 548
Db 487 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 506
QY 549 ACATTGAGACGACATCAGATGTTCCCTTCAGAAATCAAAACAAAGAGGTTGAAGAAAT 608
Db 507 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys***ValGluGluAsn 526
QY 609 TCTTTGGGATTTCTGAGAGTCTCCGTGAGACTGTTTCACAGAGGATGTGTGTACCCCAAG 668
Db 527 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 546
QY 669 GCTACACATCAAAAGAAATGGATAAAATAAGTGGAAATTAAGAGATTCACATAGCCTTA 728
Db 547 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 566
QY 729 TCAAAATCTTGATACACTTCACTCTTGTGAAGAGCAAGGAACTTCAAAAGATCAC 788
Db 567 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 586
QY 789 TGTGAACAAACGTACAGAAAAATGGAACAAATGAAAAAGAGTTTTTGTGTACTGAAAAAG 848

Db 587 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 606
 QY 849 AAACCTGTGAGAGCAAAAGAAATATAATTCAGTTAGAGAACCAAAAGAAAGTTAAATCGGAA 908
 Db 607 LysLeuSerGluAlaLysGluLysSerGlnLeuGluAsnGlnLysValTyrGlu 626
 QY 909 CAAGAGCTCTCAGTGTGAGATTGACCTTAAACCAAGAGAGAGAGAGAAATGCG 968
 Db 627 GlnGluLeuLysSerValArgLeuThrLeuAsnGlnGluGluGluLysArgArgAla 646
 QY 969 GATATATTAAATGAAATATTAGGCAAGAAATAGCAAGAAATCGAAGAGAGAGAGAGAA 1028
 Db 647 AspileLeuAsnGluLysLeuArgGluGluLysArgGluGluGlnHisArgLys 666
 QY 1029 GAGTTAGAGTGAACACACTTGAACAGAGCTCTCAGATACAGATACAGATAGAG 1088
 Db 667 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgGlnAspileGluLys 686
 QY 1089 AGTGTAGAAAGTAAATTTGAATCAGGTTTCTCAGCTCATGAAATGAAATTTATCTTTA 1148
 Db 687 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 706
 QY 1149 CATGAATTTTCATGTTGAAAGGAAATTCGCAATGCTTAACTGAACTGAAATAGCCACTG 1208
 Db 707 HisGluAsnGlnMetLysLysGluLeuAlaMetLysLeuGluLeuAlaThrLeu 726
 QY 1209 AAACACCAATACAGGAAAGGAAATATAATCTTTGAGGACATTAAGATTTTAAAGAA 1268
 Db 727 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspileLysLeuLysGlu 746
 QY 1269 AAGATGCTGACATTCAGACACCTTAACTGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1328
 Db 747 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluLeuSerLeuThrLysArgAla 766
 QY 1329 TCTCAATATAGTGGGAGCTTAAAGTTCTGATAGCTGAGAACCAATGCTCACTTCTAA 1388
 Db 767 SerGlnTyrSerGlyGlnLeuLysValLeuLeuAlaGluAsnThrMetLeuThrSerLys 786
 QY 1389 TTGAGGAAACACAGCAAGAGAAATCTAGAGGCAAGAAATGAAATCAACCACTCTAGA 1448
 Db 787 LeuLysGluLysGlnAspLysGluLeuGluAlaGluLeuGluSerHisSerProArg 806
 QY 1449 CTGGCTTCTGCTGACAGACCATGATCAAAATGTGACATCAAGAGAGAGAGAGAGAG 1508
 Db 807 LeuAlaSerAlaValGlnAspHisAspGlnLeuValThrSerArgLysSerGlnGluPro 826
 QY 1509 GCTTCCACATTCAGAGAGAGCTGTTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1568
 Db 827 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 846
 QY 1569 ACCGATATATCAATGAGGTGCTCCATCAACCACTTCTGAGCTCAAGAGAGAGAGAG 1628
 Db 847 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 866
 QY 1629 AAGCTTAAATTAATCTCAATATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1688
 Db 866 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuAspGluAsnThrLeuValSer 886
 QY 1689 AGGAACATGCAAG 1748
 Db 886 r-GluHisAlaGlnArgAspGluArgGluThrGlnCysGlnMetLysGluAlaGluHis 906
 QY 1749 TGTWTCAAGCAAG 1808
 Db 906 eTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGlnGlnGlnGlnGlnGlnGln 926
 QY 1809 AGAAATATTTCAACTCAAG 1868
 Db 926 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnGlnGlnGlnGln 946
 QY 1869 ATAANGAAGCTGACAAAG 1928
 Db 946 is-LysLysAlaAspAsnLysSerLysIleThrIleAspileHisPheLeuGluArgLys 965

QY 1929 ATGCN-CATCATCTTCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1987
 Db 966 MetGlnHisLeuLeuLysGluLysAsnGluLysIlePheAsnTyrAsnAsnHisLeu 985
 QY 1988 AAAACCCGATATTTCAATATATGAAAAAATAAAAAAATAAAAAA 2027
 Db 986 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 998
 RESULT 10
 US-09-551-621-475
 ; Sequence 475, Application US/09551621
 ; Publication No. US20030104366A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Yuqin, Jiang
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Harlocker, Susan L.
 ; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
 ; FILE REFERENCE: 210121.470C5
 ; CURRENT APPLICATION NUMBER: US/09/551,621
 ; CURRENT FILING DATE: 2000-04-17
 ; NUMBER OF SEQ ID NOS: 479
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 475
 ; LENGTH: 1002
 ; TYPE: PRT
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)...(1002)
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 US-09-551-621-475
 Alignment Scores:
 Pred. No.: 2,78e-232 Length: 1002
 Score: 3261.00 Matches: 654
 Percent Similarity: 97.19% Conservatives: 3
 Best Local Similarity: 96.75% Mismatches: 15
 Query Match: 95.83% Indels: 5
 DB: Gaps: 0
 US-09-602-362B-15 (1-2030) x US-09-551-621-475 (1-1002)
 QY 9 GTTAAAGATGCTTCTGAGGCTAACTGCGGAATGAAAGTTTCTTATCCAACTAAAGCC 68
 Db 327 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 346
 QY 69 TTAGAATTGATGACATGCAAACTTTCAAAGCAGAGCTCCCGAAGAGCCATTCGCTTC 128
 Db 347 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 366
 QY 129 GAGCTGCGCATTCGAAATGCAAAAGCTGTTCCTCAATAAAGCTTGGATTTGAAGATCAA 188
 Db 367 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 386
 QY 189 CAAACATTCAGAGCAGATCAGATCTCCCATCAGAAATCCAAACAAAGAGATATGAAGAA 248
 Db 387 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 406
 QY 249 AGTTCTTGGATTCTGAGAGTCTGTGAGACTGTTCACAGAGAGAGTGTGTGTTTACCC 308
 Db 407 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 426
 QY 309 AAGCTTACATCAAAAGAGAGAGATAGATAAATAAATGAAATTTAGAGAGTCTCTGAT 368
 Db 427 LysAla***HisGlnLysGluIleAspLysLysLysLysLysLysLysLysLysLys 446
 QY 369 AATGATGTTTCTGAGGCTCCCTGAGAGTGAAGTTTCTTATCCAACTAAAGAGCTTAA 428
 Db 447 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 466

QY 429 GAATTGATGACATGCAAACTTTCAAGCAGAGACCTCCGAGAGACCATCTGCCTTCGAG 488
 Db 467 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 486
 QY 489 CCTGCCAATGAAATGCAAAAGTCTGTTCCAAATAAAGCCCTTGGAAATGGAAGAAATCAACAA 548
 Db 487 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 506
 QY 549 ACATTGAGACAGATCAGTGTCCCTTCCAGATCAAAACAAAGAGAGCTTCAAGAAAT 608
 Db 507 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys***ValGluGluAsn 526
 QY 609 TCTTGGGATCTGAGACTCTCCGTGAGACTGTTTTCAGAGAGATGTGTGTACCCCAAG 668
 Db 527 SerTrpAspSerGluSerLeuAsgluThrValSerGlnLysAspValCysValProLys 546
 QY 669 GCTACATCAAAAGAAATGGAATAAATGAAGTGAATATGAAGATCAACTGACCTA 728
 Db 547 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 566
 QY 729 TCAAAATCTTGGATACAGTTTCACTTCTGTGAAGAGCAAGGGAATCTCAAAAGATCAC 788
 Db 567 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 586
 QY 789 TGTGAACACGCTACAGAAATCGACAAATGAAAGAGAGTGTGTCTACTGAAAG 848
 Db 587 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 606
 QY 849 AAAGTGTGAGAAGCAAAAGAAATGAAATCACTTAGAGAACCAAAAGCTTAAATGGGAA 908
 Db 607 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 626
 QY 909 CAAGAGCTCTGAGTGTGAGATTCAGTTTAAACCAAGAGAGAGAGAGAGAAATGCC 968
 Db 627 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGlnLysArgAsnAla 646
 QY 969 GATATATTAAATGAAAAATAGGGAAGATTAGGAAGATTCGAGAGCAGCATAGGAAA 1028
 Db 647 AspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGlnHisArgLys 666
 QY 1029 GAGTTAGAGTGAACAAACACTTGAACAGCTTCAGATACAGATATAGAAATGAAG 1088
 Db 667 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 686
 QY 1089 AGTGAGAAATGATTTGAATCAGGTTTCTCACACTCATGAAATGAAATATCTCTTA 1148
 Db 687 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 706
 QY 1149 CATGAAATTCATGTTGAAAGAGAAATGCCATGCTAAAGTGAAGATAGCCACTG 1208
 Db 707 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 726
 QY 1209 AAACACCAATACCAAGAAAGAAATATAATCTTTGAGGACATTAAAGATTTTAAAGAA 1268
 Db 727 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 746
 QY 1269 AAGAAATCTGAACTTCAGATGACCTTAAATCAAGAGAGAAATCAATTAAGAGGCA 1328
 Db 747 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 766
 QY 1329 TCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGAACCAATGTCTCACTTCTAAA 1388
 Db 767 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 786
 QY 1389 TTGAAGGMAAACAGCAAGAAATATCTAGAGCAGAAATTCGAATCACACCTCTCTAGA 1448
 Db 787 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 806
 QY 1449 CTGGCTTCTGTGTACAGAGCAATGATCAAAATGTGCATCAAGAAAAAGTCAAGAACT 1508
 Db 807 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 826

QY 1509 GCTTTCCACATTCAGAGATGCTTTGTTGCAAGAAAAATGAATGTTGATGTGAGTAGT 1568
 Db 827 AlaphHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 846
 QY 1569 ACCGATATATAACAATGAGGTGCTCATCAACACACTTTTGAAGCTCAAGGAAATCCAN 1628
 Db 847 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLy 866
 QY 1629 AAGCCTAAAAATTAATCTCAATTTATCAGAGATGCTCTAAGAGAAAAATACATTTGTTTC 1688
 Db 866 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 886
 QY 1689 AGGAACATGCACAAAAGAGACCAACGTGAACACAGTGTCAAATGAAGAGCTGAACACA 1748
 Db 886 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 906
 QY 1749 TGTNTCAAACNCGAACCAAGATNATGTCAACAAACACACTGANCAGCAGGAGTCTCTAGATC 1808
 Db 906 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 926
 QY 1809 AGAAATTTTCAACTACAAAGCAAAATATGTGGCTTCAACAGCAATTAGTTTCATGCAC 1868
 Db 926 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 946
 QY 1869 ATAANGAAAGCTCACAAACAAAGCAAGATAACAATTTGATNTTCATTNTCTTCAGAGGAAA 1928
 Db 946 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 965
 QY 1929 ATGCN-CATCATCTTTAAAGAGAAAAATGAGGAGATATTTNATTACNATTAACCATTTA 1987
 Db 966 MetGlnHisHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 985
 QY 1988 AAAAAACCGTATATTTCAATATCGAAAAAANAANAANAANAANAANAANAANAANAANA 2027
 Db 986 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 998

RESULT 11

US-10-007-805-475
 ; Sequence 475, Application US/10007805
 ; Publication No. US20020150581A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Yugui
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Hepler, William T.
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: Fanger, Gary R.
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: McNeill, Patricia D.
 ; APPLICANT: Durham, Margaria
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; FILE REFERENCE: 210121.470C10
 ; CURRENT APPLICATION NUMBER: US/10/007,805
 ; CURRENT FILING DATE: 2001-12-07
 ; NUMBER OF SEQ ID NOS: 593
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 475
 ; LENGTH: 1002
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: 310, 429, 522
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 US-10-007-805-475

Alignment Scores:
 Pred. No.: 2,78e-232 Length: 1002
 Score: 3261.00 Matches: 654
 Percent Similarity: 97.19% Conservative: 3

Best Local Similarity: 96.75% Mismatches: 15
Query Match: 95.83% Indels: 5
DB: 13 Gaps: 0

US-09-602-362E-15 (1-2030) x US-10-007-805-475 (1-1002)

```
QY 9 GTTAAAGTGTCTCTCAAGGCTTAACCTGCGGATGAAGTTCTATTCTCAACTAAGCC 68
DB |||||
DB 327 VallyAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 346
QY 69 TTAGAATTGATGGACATCAAACTTTCAGACGAGGCTCCCGAGAGCCATCTGCCTTC 128
DB |||||
DB 347 LeuGluLeuMetAspMetGlnThrPhelLysAlaGluProGluLysProSerAlaPhe 366
QY 129 GAGCTGCGCATTAAGTAAAGTCTCTCCCAATAAAGCTTGGAAATGAAGATGA 188
DB |||||
DB 367 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 386
QY 189 CAACATTGAGACGATGAGATCTCCCATCAAGATCCAAACAAAGGACTATGAAGAA 248
DB |||||
DB 387 GlnThrLeuArgAlaAspGluLeuProSerGluSerLysGlnLysAspFyrGluGlu 406
QY 249 AGTTCTTGGATCTGAGAGTCTCTGAGACTGTTTCAGAGAGGATGCTGTTTACCC 308
DB |||||
DB 407 SerSerIlePaspSerGluSerLeuLysGlnLysValSerGlnLysAspValCysLeuPro 426
QY 309 AAGCTCATCATCAAAAGAAATAGATAAAATAAATGAAATTTAGAAAGTCTCTCTGAT 368
DB |||||
DB 427 LysAla**HisGlnLysGluLeuAspLysIleAsnGlyLysLeuGluLysProAsp 446
QY 369 AATGATGTTTCTGAAGCTCCCTGAGATGAAGTTCTATTCTCAACTAAGCCTTA 428
DB |||||
DB 447 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 466
QY 429 GAATTGATGGATCGCAAACTTCAAGACAGAGCTCCCGAGAGCCATCTGCCTTCGAG 488
DB |||||
DB 467 GluLeuMetAspMetGlnThrPhelLysAlaGluProGluLysProSerAlaPheGlu 486
QY 489 CTTGCTGATTAAGTCAAAAGTCTTCCCAATAAAGCTTGAATTTGAAGATGAACAA 548
DB |||||
DB 487 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLysAsnGluGln 506
QY 549 ACATTGAGACGATCGATGTTCCCTCAAGATCAAAACAAAGAGGTTGAAGAAAT 608
DB |||||
DB 507 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys**ValGluGluAsn 526
QY 609 TCTTGGGATCTGAGAGTCTCGTGAGACTGTTTCAGAGAGGATGTTGTGTACCCAG 668
DB |||||
DB 527 SerIlePaspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 546
QY 669 GCTACATCAAAAGAAATCGATAAATAAGTGAAGAAATTAGAAGATCAACTAGCCTA 728
DB |||||
DB 547 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 566
QY 729 TCAAAATCTTGGATACAGTTCATTCTGTGAAGACGACAGGCACTTCAAAAAGATCAC 788
DB |||||
DB 567 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 586
QY 789 TGTGAACACCTGACAGAAATCGAACAATGAAAGAAAGTTTGTGTACTGAAAGAA 848
DB |||||
DB 587 CysGluGlnArgThrGlyLysMetGluMetLysLysLysPheCysValLeuLysLys 606
QY 849 AAATCTGTCAGAGCAAAAGAAATCAAGTTACAGTTCAGAGAACCAAAAGTTAAATGGGAA 908
DB |||||
DB 607 LysLeuSerGluAlaLysGluLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 626
QY 909 CAAGAGCTCTCGAGTGTGAGATTGACTTTTAAACAAGAGAGAGAGAGAGAAATGCC 968
DB |||||
DB 627 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgAsnAla 646
QY 969 GATATTTAAATGAAATTTAGGAGAAATTTAGGAGAAATTCAGAGAGAGAGAGAGAA 1028
DB |||||
DB 647 AspIleLeuAsnGluLysIleArgGluLeuGlyArgIleGluGlnHisArgLys 666
```

```
QY 1029 GAGTTGAAGTGAACAAACAACTTGAACAGGCTCTCAGAAATCAAGATATAGAAATGAAG 1088
DB |||||
DB 667 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 686
QY 1089 AGTGTCAAAAGTAAATTTGAATCAGGTTCTTCACACTCATGAAATGAAATATCTCTTA 1148
DB |||||
DB 687 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTrpLeuLeu 706
QY 1149 CATCAAAATTCGATGTGTAAGAAAGAAATTCGATCTAAACTGAAATGCAACTG 1208
DB |||||
DB 707 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 726
QY 1209 AAACACCAATACAGAGAAAGAAATAAATACTTTGAGCAGCATTAAGATTTTAAAGAA 1268
DB |||||
DB 727 LysHisGlnTrpGlnGluLysGluAsnLysTrpPheGluAspIleLysIleLeuLysGlu 746
QY 1269 AAGAACTCTCAACTTCAGATGACCTTAAACTGAAAGAGAAATCAATTAACCTAAAGGCA 1328
DB |||||
DB 747 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 766
QY 1329 TCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGAACACAACTGCTCACTCTAAA 1388
DB |||||
DB 767 SerGlnTrpSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 786
QY 1389 TTGAAGGAAACAAAGCAAAAGAAATCTAGAGGCGAGAAATTTGAATCAACATCTCTAGA 1448
DB |||||
DB 787 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 806
QY 1449 CTGCTTCTCTCTGTACAGACCATGATCAAAATTTGTGACATCAAGAAAGTCAAGAACT 1508
DB |||||
DB 807 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 826
QY 1509 GCTTTCACATTCGACGAGATGCTTGTGTTGCAAGAAATAAATGATGTTGATGAGTAGT 1568
DB |||||
DB 827 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 846
QY 1569 ACCGATATATAACAAATGAGTGTCTCCATCAACACTTCTGAAGCTCAAGGAAATCCAN 1628
DB |||||
DB 847 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 866
QY 1629 AAGCTTAAATTAATCTCAATTTATGACGAGATGCTCTAGAGAAATATCAATGCTTTC 1688
DB |||||
DB 866 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 886
QY 1689 AGGAACATGCACAAAGAGACCAACGCTGAACACAGTGTCAATGAAGGAAAGTCAACACA 1748
DB |||||
DB 886 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 906
QY 1749 TGTNTCAANCGAAACAAGATNATGTGAACAAACACACTGANCAGCAGAGTCTCTAGATC 1808
DB |||||
DB 906 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 926
QY 1809 AGAAATTTATTTCACTACAAAGCAAAATATCTGGCTTCAACAGCAATTAGTTCATGCAC 1868
DB |||||
DB 926 LysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 946
QY 1869 ATAANGAAGCTGACAAACAAAGCAAGATACAAATTTGATTTTCAATTTCTTTCAGAGGAA 1928
DB |||||
DB 946 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 965
QY 1929 ATGNC-CATCATCTCTTAAAGAGAAATAGAGAGATTTTATTTACNATTAACCATTTA 1987
DB |||||
DB 966 MetGlnHisHisLeuLeuLysGluLysAsnGluIlePheAsnTyrAsnAsnHisLeu 985
QY 1988 AAAAACCCGTATTTCAATATGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2027
DB |||||
DB 986 LysAsn-ArgIleTyrGlnTrpGluLysGluLysAlaGlu 998
```

RESULT 12
US-10-076-622-475
; Sequence 475, Application US/10076622
; Publication No. US20030023036A1

```
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C11
; CURRENT APPLICATION NUMBER: US/10/076,622
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 475
; LENGTH: 1002
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 310, 429, 522
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-076-622-475

Alignment Scores:
Pred. No.: 2,78e-232 Length: 1002
Score: 3261.00 Matches: 654
Percent Similarity: 97.19% Conservative: 3
Best Local Similarity: 96.75% Mismatches: 15
Query Match: 95.83% Indels: 5
DB: 14 Gaps: 0

US-09-602-362e-15 (1-2030) x US-10-076-622-475 (1-1002)
QY 9 GTTAAAGATGGTCTTCTGAAGGCTAACTGGGAATGAAAGTTTCTATTCCAACTAAAGCC 68
DB 327 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 346
QY 69 TTAGAATTGATGACATGCAAACTTTCAAAGCAGAGCTCCCGAAGCCATCTCCCTTC 128
DB 347 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPhe 366
QY 129 GAGCCTGCCATTGAAATGCAAAAGTCTGTCCAAATAAGCCTTCGAATGGAATGAAGATGAA 188
DB 367 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 386
QY 189 CAAACATTGAGAGCAGATGAGATCTCCATCAGAAATCCAAACAAAGAGCTATGAAGAA 248
DB 387 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 406
QY 249 AGTCTTGGGATCTCAGAGTCTCTGTGAGACTGTTTCACAGAGGATGTGTGTTACCC 308
DB 407 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 426
QY 309 AAGGCTACACATCAAAAAGAAATAGATAAAATAAATGGAATTAAGAAGATCTCCTGAT 368
DB 427 LysAla**HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 446
QY 369 AATGATGTTTCTGAGGCTCCCTGCAAGTAAAGTTTCTATTCCAACTAAAGCCTTA 428
DB 447 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 466
QY 429 GAATTGATGACATGCAAACTTTCAAAGCAGAGCTCCCGAAGCCATCTGCTTCAG 488
DB 467 GluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu 486
QY 489 CTGCGATTGAAATGCAAAAGTCTTTCAAATAAAGCCTTGGAAATTAAGAATGAACAA 548
DB 487 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 506
QY 549 ACATTGAGAGCAGATGATGCTTCCTTCGAATCAAAACAAAGAGGTTCAAGAAAT 608
DB 507 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys**ValGluGluAsn 526
QY 609 TCTTGGGATTTCTGAGATCTCCGTGAGACTGTTTCACAGAGGATGTGTGTACCCCAAG 668
DB 668
```

```
DB 527 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 546
QY 669 GCTACACATCAAAAAGAAATCGATAAATAAGTCGAAAATAGAGATTCACCTAGCCTA 728
DB 547 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerTrpSerLeu 566
QY 729 TCAAAAATCTTTGGATACAGATTTCATTTCTGTGAAAGAGCAAGGGAACCTTCAAAAAGATCAC 788
DB 567 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 586
QY 789 TGTGACACATGACAGGAAAAATGACAAATGAAAAAGAAATTTTGTGTCTCGAAAAAG 848
DB 587 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys 606
QY 849 AAACCTGTCAGAAAGCAAAAATAAATCAAGTTAGAGAACCAAAAAGTTAAATGGCAA 908
DB 607 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValTrpGlu 626
QY 909 CAAGAGCTCTCAGTGTGAGATTGACTTTAAACAAGAGAGAGAGAGAGAGAAATGCC 968
DB 627 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 646
QY 969 CATATATTAAATGAAAAATTTAGGGAAGAAATTAGGAAGAAATCGAAGACGACGATAGAAA 1028
DB 647 AspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGluGlnHisArgLys 666
QY 1029 GAGTTAGAAGTCAAAACAACTTCAACAGGCTCTCAGAATACAGATATAGATTGAAG 1088
DB 667 GluLeuGluValLysGlnGlnLeuGluAlaLeuArgIleGlnAspIleGluLeuLys 686
QY 1089 AGTCTAGAAATGATTTGAATTCAGGTTCTCACACTCATGAAATGAAAATATCTCTTA 1148
DB 687 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 706
QY 1149 CATGAAATTCGATGTTGAAAAAGGAAATTCGCAATGCTAAACTGGAATAGCCACACTG 1208
DB 707 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 726
QY 1209 AAACACCAATACCAAGGAAAAAGGAAAAATAAATACCTTTGAGACATTAAGATTAAAAAGAA 1268
DB 727 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 746
QY 1269 AAGAATCTCAACTTCAGATGACCTTAAACTAAAGAGGAAATCATTAACATAAAGGCA 1328
DB 747 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 766
QY 1329 TCTCAATATAGTGGGAGCTTAAAGTTCTGATAGTGAGAACAAATGCTCACTTCTAAA 1388
DB 767 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 786
QY 1389 TTGAAGGAAAAACAAGACAAAGAAATACTAGAGCGAGAAATTCGAATCAGACCATCTCTAGA 1448
DB 787 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisProArg 806
QY 1449 CTGCTTCTGCTGTACAAGACCATGATCAAAATGTCACATCAAGAAAAAGTCAAGAACCT 1508
DB 807 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 826
QY 1509 GCTTTCCACATTCAGGAGATGCTGTTTTCGAAGAGAAAAATGATGATGATGAGTAGT 1568
DB 827 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 846
QY 1569 ACCGATATATAACAATGAGGTGCTCCATCAACACCTTTCTGAAGCTCAAGAGAAATCCAN 1628
DB 847 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 866
QY 1629 AACCTTAAAAATTAATCTCAATTATCCAGAGATGCTCTAAGAGAAAAATACATTGTTTC 1688
DB 866 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSer 886
QY 1689 AGAAACATGCACAAAGAGACCAACGTTGAACACAGTGTCAATGAGGAGGCTGACACACA 1748
DB 886 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisn 906
```



```
Db 767 SerGlnTyr:SerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 786
QY 1389 TTGAAGCAAAACAGACAAAGAAATACCTAGAGGCGAGAAATGAATACACACCATCCTAGA 1448
Db 787 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisProArg 806
QY 1449 CTGGCTTCCTGCTGACACACATGATCAAAATTTGTGACATCAAGAAAAGTCAAGAACT 1508
Db 807 LeuAlaSerAlaValGlnAspHisaspGlnIleValThrSerArgLysSerGlnGluPro 826
QY 1509 GCTTTCCACATTCGAGGAGATCTGTGTTTGCAAGAAAATGAATGTTGATGTAGTAGT 1568
Db 827 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 846
QY 1569 ACCGATATATAACATGAGGCTGCTCCATCAACACCTTTCTGAGCTCAAGCAAAATCCAN 1628
Db 847 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLy 866
QY 1629 AAGCCTAAATAATCTCAATATTCAGAGAGATGCTCTAAGAGAAAATACATGCTTTC 1688
Db 866 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValse 886
QY 1689 AGGACATGCACAAAGAGACCAACGTGAACACAGTGTCAATGAAGAGAGCTGAACACA 1748
Db 886 r-GluHisAlaGlnArgaspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 906
QY 1749 TGTNTCAANCGAACAAAGATNATGTCAACAAACACACTGANCAGCAGAGTCTCTAGATC 1808
Db 906 etTyr-GlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 926
QY 1809 AGAAATATTCACTACAAACAAATAATATGTGCTTCAACAGCAATAGTTCATGCAC 1868
Db 926 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 946
QY 1869 ATAANGAAGCTGACAAACAAAGCAAGATAACAAATGTATNTTCATTNTCTTGAGAGAAA 1928
Db 946 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 965
QY 1929 ATGCN-CATCAPCTCTTAAAGAGAGAAAATAGGAGATATTNATACNATAACCATTTA 1987
Db 966 MetGlnHisHisLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 985
QY 1988 AAARACCCGTATATTCAATATGGAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAG 2027
Db 986 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 998

RESULT 14
US-10-007-805-493
; Sequence 493, Application US/10007805
; Publication No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Xuqiu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 493
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 403, 522, 615
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-007-805-493

Alignment Scores:
Pred. No.: 2,81e-232 Length: 1095
Score: 3261.00 Matches: 654
Percent Similarity: 97.19% Conservative: 3
Best Local Similarity: 96.75% Mismatches: 15
Query Match: 95.83% Indels: 5
DB: 13 Gaps: 0

US-09-602-362E-15 (1-2030) x US-10-007-805-493 (1-1095)
QY 9 GTTAAAGATGGTCTTCTGAAGGCTAACTGCGGAATGAAAGTTTCTATTCCAACTAAAGCC 68
Db 420 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 439
QY 69 TTGAATTTGATGACATGCAAACTTTCAAGCAGAGAGCTCCCGAGAGAGCCATCTGCCTTC 128
Db 440 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 459
QY 129 GAGCCTCCCATTTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAAATTTGAAGAATGA 188
Db 460 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 479
QY 189 CAAACATTTGAGCAGAGATGAGATACCTCCATCAGATCCAAACAAAGAGACTATGAAGAA 248
Db 480 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 499
QY 249 AGTTCTTTGGATTTCTGAGAGTCTGTGAGAGTGTTCACAGAGAGATGTGTGTTTACCC 308
Db 500 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 519
QY 309 AAGCTACACATCAAAAGAAATAGATATAAATAATGGAATAATAGAGAGTCTCTCTGAT 368
Db 520 LysAla***HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProasp 539
QY 369 AATGATGCTTTCTGAGGCTCCCTGCAGATCAAAAGTTTCTATTCCAACTAAAGCCTTA 428
Db 540 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 559
QY 429 GAATTGATGACATGCAAACTTTCAAAGCAGAGAGCTCCCGAGAGAGCCATCTGCCTTCGAG 488
Db 560 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 579
QY 489 CCTGCCATTTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGCATTTGAAGATGAACAA 548
Db 580 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 599
QY 549 ACATTGAGAGCAGATCAGATGTTCCCTTTCAGAAATCAAAACAAAGAGGTTGAAGAAAT 608
Db 600 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys***ValGluGluAsn 619
QY 609 TCTTGGATTTCTGAGAGTCTCCGTGAGACTGTTTTCACAGAGAGATGTGTGTACCCAAG 668
Db 620 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 639
QY 669 GCTACACATCAAAAGAAATGGAATAAATAAGTGGAAAATTAGAAGATTCAACTAGCCTA 728
Db 640 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 659
QY 729 TCAAAATCTTGGATTCAGTTCAATTTCTGTGAAGCAAGGAACTTCAAAAAGATCATC 788
Db 660 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 679
QY 789 TGTGAACACGTACAGGAAAATTTGGACAAATTCAAAAGAAAGTTTGTCTACTGAAAAG 848
Db 680 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys 699
QY 849 AAATCTGCAGAGCAAAAGAAATATAAATCACAGTTAGAGAACCAAAAGTTAAATGGGAA 908
```


| | | | |
|----|------|--|------|
| Qy | 489 | CCTGCCATTGAAATGCAAAAGCTGTGTTCTCCAAATAAAGCCTTGGAAATTGAAGAATGAACAA | 548 |
| Db | 580 | ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln | 599 |
| Qy | 549 | ACATTGAGACGACATCAGATGTCCTCCCTTCAGAAATCAAAACAAAAGAGAGGTTGAAGAAAT | 608 |
| Db | 600 | ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys**ValGluGluLysn | 619 |
| Qy | 609 | TCTTGGATCTCGAGATCTCCGTGAGACGTGTTTTCACAGAAGGATGTGTGTACCCCAAG | 668 |
| Db | 620 | SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys | 639 |
| Qy | 669 | GCTACACATCAAAAGAAATCGATATAAATAGTCGAAATTTAGAAGATTTCAACTAGCCTA | 728 |
| Db | 640 | AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu | 659 |
| Qy | 729 | TCAAAAATCTTTGGATACAGTTTCATTTCTTGTGAAAGACGAAAGGAACTTCAAAAAGATCAC | 788 |
| Db | 660 | SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis | 679 |
| Qy | 789 | TGTGAACACGTACAGGAAAATCGAACAAATGAAAAAGAAAGTTTTGTGTACTGAAAAAG | 848 |
| Db | 680 | CysGluGlnArgThrGlyLysMetGlnMetLysLysLysPheCysValLeuLysLys | 699 |
| Qy | 849 | AAACTGTCAAGAAGCAAAAGAAATAAATACAGTTTACAGAACCAACAAAAGTTTAAATGCGAA | 908 |
| Db | 700 | LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu | 719 |
| Qy | 909 | CAAGAGCTGTCAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAGAAATGCC | 968 |
| Db | 720 | GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgAsnAla | 739 |
| Qy | 969 | GATATATTAAATGAAAAATTTAGGGAAGAATTAGGAAGAATCGAAGAGCAGCATAGGAAA | 1028 |
| Db | 740 | AspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGluGlnHisArgLys | 759 |
| Qy | 1029 | GAGTTAGAAGTGAACCAACAACTTGAACAGGCTCTCAGAATACAGAATATAGAAATTGAAG | 1088 |
| Db | 760 | GluLeuGluValLysGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys | 779 |
| Qy | 1089 | AGTGTAGAAGTAAATTTGAATCAGGTTTCTCACACTCATGAATGAAATTTATCTCTTA | 1148 |
| Db | 780 | SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu | 799 |
| Qy | 1149 | CATCAAAATTCGATGTTGAAAAAGAAATTCGCCATGCTAAACTGGAATATAGCACACTG | 1208 |
| Db | 800 | HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu | 819 |
| Qy | 1209 | AAACACCAATACAGGAAAAGGAAAAATAATCTTTGAGACATTAAGATTTTAAAGAA | 1268 |
| Db | 820 | LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu | 839 |
| Qy | 1269 | AAGAATGCTGAACCTTCAGTCACCTTAAACTGAAAGAGAGAATCATTAACCTAAAGGGCA | 1328 |
| Db | 840 | LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla | 859 |
| Qy | 1329 | TCTCAATATAGTGGCAGCTTAAAGTTCTGTAGACTGAGAACCAATGCTCACTTCTTAA | 1388 |
| Db | 860 | SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys | 879 |
| Qy | 1389 | TTGAAGGAAAAACAAGACAAAGAATACTAGAGCAGAAATTTGAATCACACCATCTTAGA | 1448 |
| Db | 880 | LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisAspArg | 899 |
| Qy | 1449 | CTGCTCTTCTGTCACAGACCATGATCAAAATCTGTACATCAAGAAAAAGCTCAAGAACCT | 1508 |
| Db | 900 | LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro | 919 |
| Qy | 1509 | GCTTTCCACATTCGAGAGATGCTTGTGTTGCAAGAAAAATGAATGTGTAGTAGTAGT | 1568 |
| Db | 920 | AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer | 939 |
| Qy | 1569 | ACCGATATATAACAAATGAGGTGCTCCATCAACCACTTCTGAAGCTCAAAAGGAAATCCAN | 1628 |

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 08:46:47 ; Search time 193.218 Seconds
(without alignments)
134339.296 Million cell updates/sec

Title: US-09-602-362E-22

Perfect score: 7086

Sequence: 1 ctactgtatatacagcaacgac.....gagtggaactccacttgaaa 4115

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO spoop/US09602362/runat_15072004_093626_22001/app_query.fasta_1.10325
-DB=SPTREMBL_25 -QFMT=fastan -SUFFIX=sept -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=sept -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09602362@cgn 1 1 8.4 @runat_15072004_093626_22001 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_muc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_xodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|--------|--------------------|
| 1 | 1234 | 17.4 | 1341 | 4 | Q9BXX3 | Q9bxx3 homo sapien |

| ID | Q9BXX3 | PRELIMINARY; | PRT; | 1341 AA. |
|----|---|--------------|------|----------|
| AC | Q9BXX3; | | | |
| DT | 01-JUN-2001 (TrEMBLrel. 17, Created) | | | |
| DT | 01-JUN-2001 (TrEMBLrel. 17, Last sequence update) | | | |
| DT | 01-OCT-2003 (TrEMBLrel. 25, Last annotation update) | | | |
| DE | Breast cancer antigen NY-BR-1. | | | |
| OS | Homo sapiens (Human) | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | |
| OX | NCBI_TaxID=9606; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE=21174979; PubMed=11280766; | | | |
| RA | Jaeger D., Stockert E., Guere A.O., Scanlan M.J., Karbach J., | | | |
| RA | Jaeger E., Knuch A., Old L.J., Chen Y.F.; | | | |
| RT | "Identification of a Tissue-specific Putative Transcription Factor in | | | |
| RT | Breast Tissue by Serological Screening of a Breast Cancer Library.;" | | | |
| RL | Cancer Res. 61:2055-2061(2001). | | | |
| DR | EMBL: AF269087; AAK27325.1; - | | | |
| DR | GO: 0005634; C.nucleus; NAS. | | | |
| DR | GO: 0005515; F.protein binding; NAS. | | | |
| DR | GO: 0003700; P.transcription factor activity; NAS. | | | |
| DR | GO: 0006355; P.regulation of transcription, DNA-dependent; NAS. | | | |

ALIGNMENTS

RESULT 1

| ID | Q9BXX3 | PRELIMINARY; | PRT; | 1341 AA. |
|----|---|--------------|------|----------|
| AC | Q9BXX3; | | | |
| DT | 01-JUN-2001 (TrEMBLrel. 17, Created) | | | |
| DT | 01-JUN-2001 (TrEMBLrel. 17, Last sequence update) | | | |
| DT | 01-OCT-2003 (TrEMBLrel. 25, Last annotation update) | | | |
| DE | Breast cancer antigen NY-BR-1. | | | |
| OS | Homo sapiens (Human) | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | |
| OX | NCBI_TaxID=9606; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE=21174979; PubMed=11280766; | | | |
| RA | Jaeger D., Stockert E., Guere A.O., Scanlan M.J., Karbach J., | | | |
| RA | Jaeger E., Knuch A., Old L.J., Chen Y.F.; | | | |
| RT | "Identification of a Tissue-specific Putative Transcription Factor in | | | |
| RT | Breast Tissue by Serological Screening of a Breast Cancer Library.;" | | | |
| RL | Cancer Res. 61:2055-2061(2001). | | | |
| DR | EMBL: AF269087; AAK27325.1; - | | | |
| DR | GO: 0005634; C.nucleus; NAS. | | | |
| DR | GO: 0005515; F.protein binding; NAS. | | | |
| DR | GO: 0003700; P.transcription factor activity; NAS. | | | |
| DR | GO: 0006355; P.regulation of transcription, DNA-dependent; NAS. | | | |

Db 607 ----GlnThrTrpArgAlaAsp-GluIleLeuProSerGluSerLysGlnLysAspTyrG 625
 QY 1823 GAAAA---TTCTGGGATCTGGAGTCTCTG---TAGACTGTTTCCACAGAGGATG-TGT 1875
 Db 625 LuGluAsnSerTrpAspThr-GluSerLeuGluThrValSerGlnLysAspValCys 644
 QY 1876 TTACCAAGG---CTCGCTCAAAAGAAATAGATAAAATAATGGAATAATTAGAGGGT--- 1929
 Db 645 LeuProLysAlaAlaHisGlnLysGluIleAspLysIleAsnGlnLysLeuGluGlySer 664
 QY 1930 CCTGTAAAGTGTCTCTGAGGCTAACT---GGGAATGAAAGTTCTATTCCAACTAAA 1986
 Db 665 ProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLys 684
 QY 1987 GCCTAGA---ATTGATGATGCACTCAAACTTCAAAG---CAGSCCTCCCGAGAGCATCTGC 2040
 Db 685 Ala-LeuGluLeuMetAspMetGlnThrPhenylAlaGluProGluLysProSerAl 704
 QY 2041 CTTGAGGCTCATTGAAA---TCAAAAGTGTCTCCAAATAAAC---CTTGGAAATTAAGA 2094
 Db 704 aPheGluProAlaIleGluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLys 724
 QY 2095 ATGAACAAACATTGAGGCGAGATGAGAA---CTCCCATCA---AATCCAAACAAGGACTA 2148
 Db 724 snGluGlnThrLeuArgAlaAsp-GluIleLeuProSerGluSerLysGlnLysAspTyr 743
 QY 2149 TGAGAAAGTCTTGGGATTCGAGTCTCTG---TAGACTGTTTCCAGAGAGGATGTG--- 2203
 Db 744 GluGluSerSerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCys 763
 QY 2204 -TTTACCAGGCTCACATCAACAAAGAAATAGATAAAA---TAATGGAAAATAGAGATC 2259
 Db 763 sLeuProLysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGlu 783
 QY 2260 TCT-GAATAAGAGGTTTCTGAAGCTCCCT---GAGAATGAAAGTTCTATTCCACTAAA 2315
 Db 783 rProAspAsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLys 803
 QY 2316 GCCT---TGAATTGATGCATGCAAACTTCCAAAGCGGCTCCG---AGAGCCATCTG 2369
 Db 803 sAlaLeuGluLeuMetAspMetGlnThrPhenylAlaGluProGluLysProSerAl 823
 QY 2370 CTTGAGGCTCATTGAAAT---GAAAAGTCTGTTCCAAATAAAC---CTTGGAAATTAAG 2423
 Db 823 aPheGluProAlaIle-GluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLys 842
 QY 2424 AATGAAACAACTTGAGACATCATGAT-----TTCCCTTCAGATCAAAACAAAGAGG 2477
 Db 843 AsnGluGlnThr---LeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLys 861
 QY 2478 TTGAGAAA---TTCTGGGATTCCTCAAGTCT---CCGTGGACTGTTTTCAGAGAGGATGG 2531
 Db 862 ValGluGluAsnSerTrpAspSerGlu-SerLeuArgGluThrValSerGlnLysAsp-V 881
 QY 2532 TGTGTACC---CAGGCTACACAAAAGAAAGAGGATAAAT---AGTGGAAATTAAGAG 2585
 Db 881 aCysValProLysAlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGlu 901
 QY 2586 ATTCACTAGCCTATCAAAAT---CTTGGTACAGTTTCAT---CTTGTGAAAGGCAAGGAA 2639
 Db 901 spSerThrSerLeuSerLysIleLeuAspThrValHisSerCysGluArgAlaArgGlu 921
 QY 2640 CTTCAAAAGACACTCTGAAACA-CGTACAGAGAAATCGAACAT---GAAAAAGACTTTT 2695
 Db 921 euGlnLysAspHisCysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheC 941
 QY 2696 GTGTACTGAAAAAGAAC---TGTCAAGACAAAGAAATAAATCACAGTAGAGAA---CCA 2749
 Db 941 ysValLeuLysLysLysLeuSerGluAlaLysGluIleLysSerGln-LeuGluAsnGln 960
 QY 2750 AAAGTTAAATGGGAACAAGAG---TCTGCAGTGTAGATTGACTTAACCAA---GAAAG 2803
 Db 961 LysValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGlu 980

QY 2804 AGAAGAGAGAAATCCCGATATATTATAAGAAAAA-----TTAGCAAGAAATTAG 2851
 Db 981 LysArg-ArgAsnAlaAspIleLeuAsnGluLysIleArgGluLeuGlyArgIle-G 1000
 QY 2852 AAGATCTGAGAGCAGCATAGAAAAGAGTTAGAAAGTGAACACACAACTTCAAG---GCTCTC 2908
 Db 1000 LuGluGlnHisArg-----LysGluLeuGluValLysGlnGlnLeu-GluGlnAlaLeu 1017
 QY 2909 AGATACAAG---ATAAGAAATTGAAGGTGTAGAAAGTAATTGAATAGAGTTTCTCA---AC 2962
 Db 1018 ArgIleGlnAspIleGluLeuLysSerValGluSerAsnLeuAsnGln-ValSerHisTh 1037
 QY 2963 TCATCAAAATGAAA---ATTATTCTTACATGAATTTGATGTTGAAAAGGAATTGCCA--- 3017
 Db 1037 rHisGluAsnGluAsnTrpLeuLeuHisGluAsnCysMetLeuLysLysGluIleAlaMe 1057
 QY 3018 -TGCAAACTCGAATAGCCACACTA---AACACCAATCCAGGAAAAAGAAAAATAATTC-T 3072
 Db 1057 tLeuLysLeuGluIleAlaThrLeuLysHis-GlnTrpGlnGluLysGluAsnLysTrp 1077
 QY 3073 TTGAGGACTTAAGATTTTAAAGA---AAAGATGCTGAACCTTAG-ATGACCCCTAAACCTGA 3128
 Db 1077 heGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuGlnMetThrLeuLysLeu 1097
 QY 3129 AAAGGAATCATTCTAAAG---CGATCTCAATATGTGGCAGCT---AAAAGTTCTGATA 3182
 Db 1097 ysGluGluSerLeuThrLysArgAlaSerGlnTrpSerGly-GlnLeuLysValLeuIle 1116
 QY 3183 CTTGAGAAACAATCTCATCTCTAAATTTGAGGAAAAACAGACAAAGAAAA-----CTAGA 3236
 Db 1117 AlaGlu-AsnThrMetLeuThrSerLysLeuLysGluLysGlnAspLysGluLeuGlu 1136
 QY 3237 GGCAGAAATTAATAA-CACCATCTCTGACTGCTGCTGCTGCTAGCAAGC-----CA 3283
 Db 1136 uAlaGluIleGluSerHisPro-----ArgLeuAlaSerAlaValGlnAspHis 1153
 QY 3284 TGATCAATTTGACATCAAGAAAAGTAAGA---ACCTGCTTCCACATTCAG---GAGAT 3337
 Db 1153 sAspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl 1173
 QY 3338 GCTGTTTCCAAAAGAAATGAATGTTGATGTG---AGAGTACGATATTAAACAATGAGTGT 3394
 Db 1173 aCysLeuGlnArgLysMetAsnValAspValSerSerThrIleTrp-AsnAsnGluVal 1193
 QY 3395 CCATCACCACCTTTC---TGAGCTCAAGGAAATCCAAAACCTAAAATATTCT---CAAT 3448
 Db 1193 euHisGlnProLeuSerGluAlaGlnArgLysSerLysLeuLysIleAsnLeuAsnTr 1213
 QY 3449 TAGCAGGAGATGCTTAAGAGA---AATACATTTGTTTTCAGAACATCACAAAGAGA---CA 3502
 Db 1213 yrAlaGlyAsp-AlaLeuArgGluAsnThrLeuValSerGluHis-AlaGlnArgAspG 1232
 QY 3503 ACCTGMAAACAGTGTCAAT---GAAGGAGCGACACATGATCAAAAC-----GACAA 3553
 Db 1232 nArgGluThrGlnCysGlnMetLysGluAlaGluHisMetTrpGlnAsnGluGlnAspAs 1252
 QY 3554 GATAATGMAACAAACCTCGAACAGCAGAGTCTCTAGTCAGAAATATTATTCAACTACAAG 3613
 Db 1252 nValAsnLysHisThrGluGlnGlnGluSerLeu-AspGlnLysLeuPheGlnLeuGln 1272
 QY 3614 CAAAAATTGTGGCTTTCACAGCAAT---AGTCATCGACATAGAAAGCTCACACAAAGAG 3669
 Db 1272 erLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaHisLys-----LysA 1288
 QY 3670 CAGATAACAAT-----GATATTCTTTCTTTG---AGAGGAAATGCAACAC 3711
 Db 1288 laAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLysMetGlnHis 1308
 QY 3712 ATCTCTTAAGAGAAAAAGAGGAGATATT---AATTACATACCATTTAAACACCGTAT 3768
 Db 1308 isLeuLeuLysGluLysAsnGlu-GluIlePheAsnTrpAsnAsnHisLeuLysAsnArg 1327

QY 3769 ATT---CAATATGAAAGAGAA 3789
 Db 1328 lIeTyrGlnTyrGluLysGluLys 1335

RESULT 2
 Q9BXX2
 ID Q9BXX2 PRELIMINARY; PRT; 1011 AA.
 AC Q9BXX2;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DE Breast cancer antigen NY-BR-1.1 (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21174979; PubMed=11280766;
 RA Jaeger D., Stockert E., Guere A.O., Scanlan M.J., Karchach J.,
 RA Jaeger E., Knuth A., Old L.J., Chen Y.T.;
 RT "Identification of a Tissue-Specific Putative Transcription Factor in
 RT Breast Tissue by Serological Screening of a Breast Cancer Library."
 RL Cancer Res. 61:2055-2061 (2001).
 DR EMBL; AF269088; AAK27326.1; -.
 DR InterPro; IPR002110; ANK.
 DR Pfam; PF00023; ank; 3.
 DR SMART; SM00248; ANK; 3.
 DR PROSITE; PS50088; ANK_REPEAT; 1.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 KW ANK repeat; Repeat.
 FT NON TER 1011 1011
 SQ SEQUENCE 1011 AA; 114250 MW; 16627D7B218DB438 CRC64;

Alignment Scores:
 Pred. No.: 2,34e-38 Length: 1011
 Score: 481.00 Matches: 363
 Percent Similarity: 44.39% Conservative: 135
 Best Local Similarity: 32.35% Mismatches: 303
 Query Match: 6.79% Indels: 327
 DB: Gaps: 75

US-09-602-362E-22 (1-4115) x Q9BXX2 (1-1011)

QY 380 CATGGTCAGTATCGAAGTGCRAA---CAAGGCTGCCTCACACCTTTTACTACCAT-- 434
 Db 8 TyrGlyAlaValIleGluValGlnAsnLysAlaSerLeuThrProLeuLeuLeu-Ala11 27

QY 435 -AACGAAAGAGTGCAGCAATTGGATT---GCTGATAAAATGCAATCGAATGC 490
 Db 27 eGlnLysArgSerLysGlnThrValGluPheLeuLeuThrLysAlaAsnAlaAsnAl 47

QY 491 GTTA---ATAAGTTAAATGCACACCTCATGCTGCTGATG---TCTGGATCATCAGAGA 544
 Db 47 aPheAsnGluSer-LysCysThrAlaLeuMetLeuAlaIleCysGluGlySerSerGluI 67

QY 545 TAGTGTGATCTCTTACGAAA---TGTCAGCTCTTTGTCAG---ATATAGTGAAGTA 598
 Db 67 leValGlyMetLeuLeuGlnGlnAsnValAspValPheAlaGluAspIle-HisGlyIle 86

QY 599 ACTGCGAAA---CATATGCTGTGTACTGTGATTTCTACATTTCATAACAAATTTATGAATA 655
 Db 87 ThrAlaGluArgTyrAlaAlaAlaArgGlyValAsnTyrIleHisGlnGlnLeuLeuGlu 106

QY 656 TAT---ACGAAATTCAAAATCATCAATACCAATCAGAGGAACCTCTCGAGGAAC 712
 Db 107 HisIleArgLysLeuProLysAsn----- 114

QY 713 CTGATGAGGCTGCACCTCGCGAAAAGACACCTGCACACCTGAAAGCTGGTGGAAAACA 772
 Db 115 -----ProGlnAsnThrAsnProGluGlyThrSerThrGly-----Thr 127

QY 773 CTTGATGAGGCTGCACCT---GGTGGAAAGACACCTGCACGCTGAAAGCT---GGTGC 826

Db 128 ProAspGluAlaAla-ProLeuAlaGluArgThrProAspThrAlaGluSerLeuLeuGlu 147
 QY 827 AAAACACACCTGATGAG---GCTGCATCTTGGTGGAGGAACATCTGACAAATTCAT--G 880
 Db 147 uLysThrProAspGluAlaAla-ArgLeuValGluGlyThrSerAlaLysIleGlnCysL 167
 QY 881 TTGGAGAAAGGACATCTGAAAAGTTGGAACGTGAGC---AGAAAAACACCTAGGAAA--- 934
 Db 167 euGlyLysAlaThrSerGlyLysPheGluGln-SerThrGluGluThrProArgLysIle 186
 QY 935 TTACGATCTCGCAAAAACATCTGAGAAATTTACGGG-CCAGCAAAAGGACACCTAGAA 993
 Db 187 LeuArgProThrLysGluThrSerGluLysPheSerTrpProAlaLysGlu----- 203
 QY 994 GATCGCAGGAGAAAAGAACACACACCTAGGAAATATGATCCCGAAAAGAACACACTG 1053
 Db 204 -----ArgSerArgLysIleThr-----TrpGluGluLysGluThrSerVal 217
 QY 1054 AGAATTCCTGGGAGAAAAGAACAGACCTAGGAGAGCGATCGGAGAAAAGAACACCT 1113
 Db 218 LysThrGluCysValAlaGlyValThrProAsnLysThrGluValLeuGluLysGlyThr 237
 QY 1114 GTAAAGATCGATCGCTGCAAGAGATAACATCTAATAAATACTAAA----- 1155
 Db 238 SerAsnMetIleAlaCysProThrLysGluThrSerThrLysAlaSerThrAsnValAsp 257
 QY 1155 ----- 1155
 Db 258 ValSerSerValGluProIlePheSerLeuPheGlyThrArgThrIleGluAsnSerGln 277
 QY 1156 -----GTTTGAAGAAGGAATC 1173
 Db 278 CysThrLysValGluGluAspPheAsnLeuAlaThrLysIleIleSerLysSerAlaAla 297
 QY 1174 TAAGATGTTGCATGT---CCACAAAAGAAATCAT-----CTACAAAGC 1212
 Db 298 GlnAsnTyrThrCysLeuProAspAlaThrTyrGlnLysAspIleLysThrIleAsnHis 317
 QY 1213 AAGTCCCATGATCAGAG---TCCCATCAGATCCAAACAGGAGATGAGATATATCT 1269
 Db 318 LysIleGluAspGlnMetPheProSerGluSerLysArgGluGluAspGluGlyThrSer 337
 QY 1270 TTGATTTCTCGGTCTCTTTGAAGTTCTGC-----AAGATTCAGTG---GTATACCTCAGT 1322
 Db 338 TrpAspSerGlySerLeuPheGluSerSerAlaLysThrGlnValCysIleProGluSer 357
 QY 1323 CTATATATAAAGTAATGATGATATAAAGAG---TAGAAGCCTCTCTAAAGCCATCT 1379
 Db 358 MetTyrGlnLysValMetGluIleAsnArgGluValGluGluLeuProGluLysProSer 377
 QY 1380 GCCTTCAAGCCGCCATGAAA---GCAAAACTCTTCCAAATAAGCCT---TTGAATGAA 1433
 Db 378 AlaPheLysProAlaValGluMetGlnLysThrValProAsnLysAlaPheGlu-LeuL 397
 QY 1434 GAATGAACAAACATTAAGACAGATCCGTGT---CCACCCGAATCCAAACAAAGG---AC 1487
 Db 397 sAsnGluGlnThrLeuArgAlaAlaGlnMetPheProSerGluSerLysGlnLysAspAs 417
 QY 1488 TATAAGAAAATTTCTGGATTCTAGATCT---CTGGAGACTGTTTACAGAGGATTTGTG 1544
 Db 417 pGlu-GluAsnSerTrpAspSer-GluSerProCysGluThrValSerGlnLysAspVal 436
 QY 1545 TTT---ACCAGAGGTACACATCAAAAAGAA---TAGATAAAAAAATGGAATAATAGAGAG 1598
 Db 437 TyrLeuProLysAlaThrHisGlnLysGluPheAspThrLeuSerGlyLysLeuGluGlu 456
 QY 1599 TCCCTATAAAGTGGTCTTCTGAGAGGCTACTCGG-----AATGAAATTTCT 1645
 Db 457 SerProValLys-----AspGlyLeuLysProThrCysGlyArgLysValSer 473
 QY 1646 ATTCCACTAAGCCT---AGAATTTGAGACATCAAACTTTCAAGCG----- 1690

Db 474 LeuProAsnLysAlaLeuGluLeuLysAspArgGluThrPheLysAlaGluSerProAsp 493
QY 1691 -----AGCCTC----- 1696
Db 494 LysAspGlyLeuLeuLysProThrCysGlyArgLysValSerLeuProAsnLysAlaLeu 513
QY 1697 -----CGGAGAGCCATCTGCTCGAGCCTCCA-----CTGAATCA 1732
Db 514 GluLeuLysAspArgGluThrLeuLysAlaGluSerProAspAsnAspGlyLeuLys 533
QY 1733 AAAGTCTCT-----CCCAATAAGCCTTGGAAATCAAAATGAAATCAAAAC 1774
Db 534 ProThrCysGlyArgLysValSerLeuProAsnLysAlaLeuGlu-----LeuLysAsp 551
QY 1775 ATGAGCAGATGAGAA-----CTCCATCAGAAATCAAAAGAGACTATAGA 1824
Db 552 ArgGluThrPheLysAlaAlaGlnMetPheProSerGluSerLysGlnLysAspAspGlu 571
QY 1825 AAA---TTCTGGTACTGGAGTCTCTAGACTGTTTCAGAGGATTG----- 1872
Db 572 GluAsnSerTrpasp-----PheGluSerPheLeuGluThrLeuLeuGlnAsn 587
QY 1873 -----TGTTTACCAAGG---CTGGCTCAAAAAGAATAGAT-----AAAATA 1911
Db 588 AspValCysLeuProLysAlaThrHisGlnLysGluPheAspThrLeuSerGlyLysLeu 607
QY 1912 ATGAAAATTAGAGGTCCTCTTAAGTGGTCTTCTAGGCTAACT---GGGAATGAA 1968
Db 608 GluGluSer-----ProAspLysAspGlyLeuLeuLysProThrCysGlyMetLys 624
QY 1969 AGTTCTATTCCTCAACTAAAGCCTCAG---ATTGATGACATGCAAACTTCAAGCAGGCCTC 2025
Db 625 IleSerLeuProAsnLysAla-LeuGluLeuLysAspArgGluThrPheLys-----Al 642
QY 2026 CCGAGAGCCATCTGCCTTCGAGCCTCCATTGGAATCAAAAGTCTGTCGAATAAACCTTG 2085
Db 642 aGluAspValSerValGluSerThrPheSerLeuPheGlyLysProThrThrGlu-- 661
QY 2086 GAATTAGAAATGAACAACATTGAGCAGAT-----GAGAACTCCCA 2127
Db 662 AsnSerGlnSerThrLysValGluGluAspPheAsnLeuThrThrLysGluGlyAlaTh 681
QY 2128 TCAATCAAAACAAAGGACTATGAGAAAGTTCTTGGATTCTGAG----- 2173
Db 681 rLysThrValThrGlyGlnGlnGluArgAspGlyIleIleGluArgAlaProGlnAs 701
QY 2173 ----- 2173
Db 701 pGlnThrAsnLysMetProThrSerGluLeuGlyArgLysGluAspThrLysSerThrSe 721
QY 2174 -----TCTCTGTAGACTGTTTCCAGAGGATGGTGTACCCAGGCTAC 2217
Db 721 rAspSerGluIleIleSerValSerAspThrGlnAsnTyrgluCysLeuProGluAlaTh 741
QY 2218 ACATCAAAAGAAATAGATAAAA-----TAATGGAAATAGAGAGTCTCTGATAATGAG 2271
Db 741 rTyrglnLysGlu-----IleLysThrThrAsnGlyLysIleGluGluSer----- 756
QY 2272 GTTTTCTGAGGCTCCCTCGAGAATCAAGATTCATTCCACTAAAGCCTTGAATTGATGCG 2331
Db 756 ----- 756
QY 2332 ATGCAAACTTCAAGCAGGCTCCCGAGACCATCTGCTTCGAGCCTGCATTGAAATGA 2391
Db 757 -----ProGluLysProSerHisPheGluProAlaThrGluMet-----G 770
QY 2392 AAAGTCTCTTCAATATA-----CCTTGGAAATTAAGATGAACAACATTGACACAGAT 2445
Db 770 nAsnSerValProAsnLysGlyLeuGluTrpLysAsnLysGlnThr----- 785
QY 2446 CAGATTCCCTTCAGAAATCAAAACAAAGGTTGAGAAAATCTGGGATTCGAGTCT 2505
Db 786 -----LeuArgAlaAspSerThrThrLeuSerLysIleLeuAspAlaLeuProS 802

QY 2506 CCGTGGACTGTTTTCAGAGAGGATGGTGTGTACCCAGGCTACACAAAAAGAGGATA 2565
Db 802 erCysGlu----- 804
QY 2566 AAATAGTGGAAAATTAGAGATTCTAGCCTATCAAAATCTGGTACAGTTTCATCTGT 2625
Db 804 ----- 804
QY 2626 GAAAGCAAGGGAACCTTCAAAAGAGACACTGTGAACACGCTAC---AGGAAAATGGAACAT 2682
Db 805 --ArgGlyArgGluLeuLysAspAsnCysGluGlnIleThrAlaLysMetGluGlnM 824
QY 2683 GAAAAGAGATTTTGTGTACT---GAAAAGAACTGTCAACAACAAAGAAATAAATCACA- 2738
Db 824 etLysAsnLysPheCysValLeuGlnLysGluSerGluAlaLysGluLysSerG 844
QY 2739 --GTAGAGAACCAAAAGTT-AAATGGGAACAAGAG---TCTGCAGTGTAGATTGACTTAA 2792
Db 844 lnLeuGlnAsnGlnLysAlaLysTrpGluGlnGluLeuCysSerValArgLeuProLeuA 864
QY 2793 ACCAA---GAAAAGAGAGAGAAATGCCGATATATTAAAGAAAATAATAGGAAT 2849
Db 864 snGlnGluGluLysArg-ArgAsnValAspIleLeuLysGluLysIle-ArgProG 883
QY 2850 AGAAGAAATCGAGCAGCATAGAAAGAGTTAGAAGTGAACACAACTTCAAAAGGCTCT-- 2907
Db 883 uGluGlnLeuArg-----LysLysLeuGluValLysHisGlnLeuGluGlnThrLe 900
QY 2908 -CAGATPACAAGAT---AGAATTGAAGGTGTAGAAAGTAATTGTAATAGTTTCTCA---- 2960
Db 900 uArgIleGlnAspIleGlu-LeuLysSerValThrSerAsnLeuAsnGln-ValSerHis 919
QY 2961 ACTCATGAATGAAA---ATTATCTTACATGAATTCGATGTGAAAAGGAATTTGCCA 3017
Db 920 ThrHisGluSerGluAsnAspLeuPheHisGluAsnCysMetLeuLysGluLysAla 939
QY 3018 ---TGCAAACTGGATAGCCACACTA---AACACCAATCCAGGAAAAGAAAATAAATC 3071
Db 940 MetLeuLysLeuGluValAlaThrLeuLysHisGlnHisGlnValLysGluAsnLysTy 959
QY 3072 TTTGAGCACTTAAGATTTTAAAGA---AAAGATGCTGAACCTTAG-ATGACCTTAAACTG 3127
Db 960 PheGluAspIleLysIleLeuGlnGluLysAsnAlaGluLeuGlnMetThrLeuLysLeu 979
QY 3128 AAAAGGAATCAATTACTATAAAG---GGATCTCAATAT 3160
Db 980 LysGlnLysThrValThrLysArgAlaSerGlnTy 991
RESULT 3
Q9NSI9 PRELIMINARY; PRT; 424 AA.
ID Q9NSI9
AC Q9NSI9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PRED4 protein (Fragment).
DE PRED4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.S.,
RA Toyoda A., Ishii K., Totoki Y., Choi D.K., Soeda E., Ohki M.,
RA Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuoka S., Antonarakis S.E.,
RA Miroshima S., Shimizu N., Nordiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,

Db 878 eluethrAsnHisLeuSerLysGlnLysGluLeuMetAlaGlnLysLysMetAsnSe 898
QY 2965 ATGAAT -----GAAATATTCTTACATGAAATTCATGTTGA 3003
Db 898 rGluAenSerHisSerHisGluGluGluLysAspLeuSerHisLys-AsnSerMetLeuG 918
QY 3004 AAAAGGAATTGCCA---TGCAAAATGAAATAGCACACTA---AACACCAATCCAGAA 3057
Db 918 InGluGluLeuAlaMetLeuArgLeuGluLeuAspThrIleLysAsnGlnAsnGlnGluL 938
QY 3058 AAGGAAATAAATCTTTGAGGACTTAAGATTTT---AAAGAAAAGATGCTCAACTTAGAT 3114
Db 938 ySGLuLysCysPheGluAspLeuLys-ileValLysGluLys-----AsnGluAsp 955
QY 3115 GACCTTAAATCTGAAAAG-----GAATCATTTCTAAAGGATCTCA---ATATGT 3162
Db 956 LeuGlnLysThrIleLysGlnAsnGluGluThrLeuThrIleSerGlnTyAsn 975
QY 3163 GGGCAGCTTAAAGTTCTGATAGTCAGAAAC---AATGCTATTCTAAATGAGGAAAAC 3219
Db 976 GlyArgLeuSerValLeuThrAlaGluAsnAlaMetLeuAsnSerLysLeuGluAsnGlu 995
QY 3220 AGACAAAGAAA-----CTAGAGGAGAAATGAATACACCATCTGCTGCTCGCT 3273
Db 996 LysGlnSerLysGluArgLeuGluAlaGluValGlu--SerTy:HisSerArgLeuAla 1015
QY 3274 GTACAGGCCATGAT-----CAATTGTGACATC 3300
Db 1015 laAlaIleHisAspArgAspGlnSerGluThrSerLysArgGluLeuGluLeuAlaPheG 1035
QY 3301 AAGAAAAGTAAGAACCTGCTTCCACATTTGAGGAGATGCTGTTTGCAGAAATGAATG 3360
Db 1035 InArgAlaArgAspGluCysSerArgLeuGlnAsp-----LysMetAsn 1050
QY 3361 TTGATCTG---AGATGAGTATTAAATGAGTCTCCATCACCATTCTGAGCTCAA 3417
Db 1050 heAspValSerAsnLysLys-AspAsnAsnGluLeu-----SerGln 1064
QY 3418 AGGAAATCCAAAACC---TAAATATTCTCAATTAGCAGAGAT----- 3459
Db 1065 GlnLeuPheLysThrGluSerLysLeuAsnSerLeuGluLeuGluPheHisThrArg 1084
QY 3460 -----GTCTAAGGAAATACATTGTTTTCAGAACATCACAAGAGAC----- 3501
Db 1085 AspAlaLeuArgGluLysThrLeuGluGluArgValGlnLysAspLeuSerGlnThr 1104
QY 3502 -----AACGTGAAACACAGTCTCAATGAGGAGCGACACATCTATCAAAACGACAGA 3555
Db 1105 GlnCysGlnMetLysGluMetGluGlnLysTyrrGlnAsnGluGlnValLysVal----- 1122
QY 3556 TAATGGAAACAAACCTGAAACAGCAGAGTCTCT---AGTCAGAAATATTTCACATACAA 3612
Db 1123 -----AsnLysTyrrIleGlyLysGlnGluSerValGluArgLeuSerGlnLeuGln 1140
QY 3613 GCAAAA---ATTGTGGCTTCACAGCATTTAGTCATGCATAGAAAGCTGACAAACAAAG 3669
Db 1141 SerGluAsnMetLeuLeuArgGlnGlnLeuAspAspAlaHisAsnLysAlaAspAsnLys 1160
QY 3670 CAGATACAATGATATTCTTTCTTGAGAG----- 3699
Db 1161 GluLysThrValIleAsnIleGlnAspGlnPheHisAlaIleValGlnLysLeuGlnAla 1180
QY 3700 -----GAAATGCAACATCTCTCTTAAAGAGAAAAGAGAGATATT-----AATTAC 3747
Db 1181 GluSerGluLysGlnSerLeuLeuGluGluArgAsnLysGluLeuIleSerGluCys 1200
QY 3748 AATACCATTTTAAAAACCGTATATTCAATATGAAAAAGAGAGAAAGAGAAACAGAACTCATG 3807
Db 1201 AsnHisLeuLysGluArgGlnTyrrGlnTyrrGlnAsnGluLysAlaGluArgGluVal 1220
QY 3808 AGAACAGCAGTAGAAATCTTTTGGAGAAAACACAGACAGACTTTTACTTCACACTCATG 3867

Db 1221 -----ValArgGlnLeuGlnGlnGlnLeuAlaAspThrLeuLysLysGlnSerMet 1237
QY 3868 CTAGAGGCCAGTCTA 3882
Db 1238 SerGluAlaSerLeu 1242
RESULT 8
Q8ILS9
ID Q8ILS9 PRELIMINARY; PRT; 3026 AA.
AC Q8ILS9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF14_0165.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Bisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum";
RL Nature 419:498-511(2002).
DR EMBL; AB014818; AAN36777.1; -.
KW Hypothetical protein.
SQ SEQUENCE 3026 AA; 357633 MW; 9ECABD915C3C25CE CRC64;
Alignment Scores:
Pred. No.: 2,018-05 Length: 3026
Score: 150.00 Matches: 255
Percent Similarity: 33.77% Conservative: 182
Best Local Similarity: 19.71% Mismatches: 433
Query Match: 2.12% Indels: 424
DB: 59 Gaps: 59
US-09-602-362E-22 (1-4115) x Q8ILS9 (1-3026)
QY 835 CTGATGAGCTGCATCTTGGTGGAGGAAACATCTGACAAATTCATGTTGGAGAAAGGACA 894
Db 1126 LeuArgLysLeuLysAsnAsnGlyAsnLysCysThrTyrrGluLysGlyGluLys 1145
QY 895 TCTGGAAAGTTTCAACGTCAGCAGAGAAAACACCTAGGAAATACGATCTGCCAAAAC 954
Db 1146 AsnLysAsnPheCysSerGlnSerArgAsnAsnSerLeuThrSerPheArgAsnLysAsn 1165
QY 955 ATCTGAGAAAT----- 966
Db 1166 AspLysAsnIleLysLysAsnValLeuThrLysLysAspSerIleLeuAsnGlnAspLys 1185
QY 967 TACGGCCAGCAAGAGAACCTAGACATCGCAGGAGAAACAGACACACCTAGCG 1026
Db 1186 IleAsnAspCysLysAspAspLysLysLysLysLysLysLysLysLysLysLysLys 1204
QY 1027 -----AAATATGATCCCGAAAGAAACACTGAGAAATTC----- 1062
Db 1205 SerGlnMetAsnGlnLysLysIleGluSerAspLysLysLysLysLysLysPheGluLys 1224
QY 1063 GTGGCCAGAAAGAGAACCTAGGAGACCGATAGGACCGATGGAGA---AAAGAAACCTGTAAG 1119
Db 1225 TyrGlyLysTyrrGlyLysAsnGluLysTyrrArgGlyLysTyrrHisLysTyrrProLeuAsp 1244
QY 1120 ATGGATGCG-----TGCAAGAGTAACTATCTAATAAAGTAAAGTTTGTGAAAAAGGAA 1170

1245 ileGluAlaSerTyrA-GCysLysAsn-LeuTyrLeuLysile-AsnValAspThrAsnA 1264
 1171 ATCTAAGATGTTGCATGTCCCAAAAGAAATCATCTACAAAGCAAGTCCCATGATCAGAGG 1230
 1264 sn-----GluGlnileLeuLysTyrile----- 1271
 1231 TCCCATCAGATCCAAACAAAGGAAGATCAAGAAATATCTTTGATTTCTCGGTCTCTTTGA 1290
 1272 -----ProAsnLys----- 1281
 1291 AGTTCTGCAAGATTCAAGTGTGTATACCTGAGTCTATATATAAAGTAAGTAAGTAATAA 1350
 1281 luPheLysLeuileleThrLysPheileAspGlyPheileLysCysThrGlu----- 1298
 1351 AGAAGTAGAAGCCCTCCTAAAAGCCATCTGCCTTCAAGCCGCCATTCGAAAGCAAACTCT 1410
 1299 -----SerLysile 1302
 1411 TTCCAAATPAAGCTTTGAATGAAGAATGAACAAACATGAGCAGATCCGTGTTCACACC 1470
 1302 yr-----LysSerAsnileleLeuGlnAsnileGluThrAsp---IleSerTyrA 1319
 1471 GAATCCAAACAAAGGACTATAGAATAATCTTTGGGATCTAGACTCTCTGGAGACTGTTT 1530
 1319 snileThrileHisAlaTyrAsnThrValValProSerLeu-----TrpLeuT 1335
 1531 ACAGAAGGATTTGTTTACCAAGGCTACATCAATCAAAAGAAATAGATATAAAATAATGGA 1590
 1335 yrAlaSerIleSerPheTyr-----LeuThrLysPheHisL 1347
 1591 TAGAAGAGTCCCTAATAAAGTGGTCTTCTGAAGGCTACCTCGGATCGAATGAAATTTCTATCC 1650
 1347 euGlu----- 1348
 1651 ACTAAAGCTAGAAATGAAGACA---TGC-----AACTTTCAAGCGAGCTCCCGGA 1701
 1349 -----AsnPhelLysLysThrProCysHisPheThrLeuAlaGluAsnThrAspPheL 1367
 1702 AGCCATCTGCTTCGAGCTCCACTGAATCAAAAGTCTGTCCCAATTAAGCTTTGGATG 1761
 1367 euAsnSerIleThrAsnAsnAspAsnGluLysLysileValProLeuLys-----AsnG 1385
 1762 AAAATGAACATGGAGGAGATGAGACTCCCATCAGAATCAAAACAAAGACTATA 1821
 1385 lulyAsnileThr----- 1390
 1822 AGAAAATTTCTGGGATACCTGGAGTCTCTGTAGACTGTTTTCAGAGGATTTGTTTACCA 1881
 1390 InLysile-----IlePheHisAsn-----IleAsnGluL 1400
 1882 AGGTCGCTCAAAAGATAGATAAAATATGAAGGATTAAGAGGTCCTGTTTAAAGT 1941
 1400 yLysLysMetLysAsnAsnGluLeuThrAsnAsnCysThr-----AsnMetLeuLysA 1418
 1942 GGTCTTCTGAGCTAACTGGGAAGTCTTATTCCAACTTAAAGCTCAAGCTGAGATGATGA 2001
 1418 spileGluGluHisAspAsnGln---HisPheTyrLysAsnAsnSer-----T 1433
 2002 CATGAACTTCAAGACGGCTCCGAGAGCCATCTGCCT---TCGAGCTCCAT--- 2054
 1433 hrCysSerAspGluAspAsnSerAspGluSerileValProLysAsnLeuHisileG 1453
 2055 -----TGAATCAAAAGTCTGTCCAAATAAACTTGGAA 2088
 1453 InAsnAsnLeuGluPheLysGlyileLysAsnGlnLysAsnGlyAsnileAsnileSerT 1473
 2089 TTAAGATGAACAACTTGAAG----- 2111
 1473 hrArgAsnAsnValThrLeuAsnAsnAsnileAsnAsnLeuAsnAsnLysTyrAsnAsnT 1493
 2112 -----CAGATGAGAACT----- 2124

1493 hrIleAspGlyLysThrGluMetAsnSerGlnileHisAsnThrIleHisGluAspGluG 1513
 2125 CCATCAAAATCCAAACAAAGGACTATGAGAAAGTCTTGGGATTTCTGGAGTCTCTGTAGAC 2184
 1513 luileLysSer---ThrSerSerTyrGluGluLeuLeuLeuSerPheHisAspLeuAsnLy 1532
 2185 TGTTTCCAGAGGATGGTGTATCCAGGCTACATCAAAAA----- 2227
 1532 sSerGluileArgAsnLysPheThrGlnValLysTyrGluAsnCysAsnAsnValAspG 1552
 2227 ----- 2227
 1552 uArgLysileThrProThrPheGluThrProSerGlyileGluTyrProSerPheCysLy 1572
 2228 -----GA 2229
 1572 sTyrAspTyrAsnSerThrSerAsnGlnLysileGluValileAsnLysValCysMetG 1592
 2230 ATAGATAAAATATGAAAAATAGAGAGTCTCTGATAATGAGGTTTCTCGAAGGCTCCCT 2289
 1592 uGluValLysAspLysLysLysGluAsnTyrileValTyrLysLysAsnLysAspAsnLy 1612
 2290 GAGAATGAAA-----GTTCTATTCCACTPAAGCCTTGAATTCATGCGCATCAAACT 2340
 1612 sGluLeuLysAsnThrThrIleLeuAsnHisAsnProArgAsnLeuAsn-----LysG 1630
 2341 TTCAAAGCAGGCTCCCGAGAGCCATCTGCTTCGAGCCTGCATTCGAAATGAAAGTCTGT 2400
 1630 nSerPheLysPheSer-GluPheSerLeuLysLysLysValAsnLeuAsnThrLysT 1650
 2401 TCCAAATAAACCTTGGAAATTAAGAATGAACAACTTGAACAGATCAGATTCCTCTTCAG 2460
 1650 hrGlnLysAsnileAsnLysAsnAsp-Asn----- 1660
 2461 AATCAAAACAAAGAGGTTGAGAAATTCGGGATTTCTGAAGTCTCCGTGAGCTGTTTCA 2520
 1661 AsnAsnAsnAsnAsnAsnAsnGluAsnAsnAspIle----- 1672
 2521 CAGAAGAGTGTGTGTACCCAGGCTACACACAAAGAAAGGATAAAATAGTGGAAATTT 2580
 1673 ---AsnAspIleAspIleSerAsnMetSerSerAsnGlu-----LeuThrile 1687
 2581 AGAAGTTCACCTAGCTATCAAAATCTTGGTACAGTTCATCTTGTGAAAGCAAGCAAC 2640
 1688 LysGluileThrAspSerAsnSerPheileLeuSerSerCysLysSerLysGluAsn 1707
 2641 TTCAAAAGACACTGTGAACACGCTACAGGAAATGCAATGCAAAAGAGATTTTGTGTA 2700
 1708 PheLysAsnAsp-----AsnLysLysLysAsnileGluLeu 1719
 2701 CTGAAAAGAAC-----TGTCAAGCAAAAGAAAT 2730
 1720 IleLysSerAsnLysAspSerThrGluAsnSerHisGlnCysGluGlu-LysAsnL 1739
 2731 AATACACAGTACAGAACCAAAAGTAAATGCGAACAGAGTCTGCAGTGTAGATTGACTTT 2790
 1739 eAsnAsnileGlnGluLysLys-----GlnThrLeuArgAspIleLeu 1754
 2791 AAACCAAGAAAAGAGAGA-----GAGAAATCCCGATATATTAAAGA 2832
 1754 uAsnGlnLeuLysGluLysAsnGlyThrTyrAspAspAspLysGlyTyrileGlnLy 1774
 2833 AAAAATTAGCAAGAA----- 2848
 1774 sLysileHisLysAsnAspAsnileAsnValAsnCysAsnileAsnLysTyrAspIleAs 1794
 2849 ----TAGAAGATTCGAGAGCAGCATAGAAGATTAGAAGTGAAGCAACACTTGAAGGC 2904
 1794 pLysAsnLysAsnileSerileAsnAsnAsnThrIleAsnAsnAsnThr----- 1811
 2905 TCTCAGATACAGATAAGATTCAGAGGTTGAGAAAGTAAATTTGAATAGTGTCTCAACTC 2964
 1811 eAsnAsnAsnThrValAsnAsnLysTyrileLysLysGluHisAsnSerPheileAsnTy 1831

QY 2965 ATGAATGAAATTAATTCCTACATGAATAATGTCATGTTGAAAAA----- 3007
 Db |||||
 QY 1831 rLysAsnAsnGluMetLysAsnIleLysLeuAspLysLysLysGlyAspGluSerIle 1851
 Db |||||
 QY 3008 -----GCAATTGCCATGCAAAATCTTGAATAGCCACACTAA 3042
 Db |||||
 QY 1851 eLysAsnTyrGluLysAsnLeuGlnAspIleTyrIleLysAsn-----AsnAsnTyrCysAs 1870
 Db |||||
 QY 3043 ACACCAATCCAGGAAAGAAATAAATCTTTGAGGACTTAAGATTTTAAAGAAAGATG 3102
 Db |||||
 QY 1870 nAsnTyrMetLysLysSerLys-----GluGluThrAspGluAsnCy 1884
 Db |||||
 QY 3103 CTGAACCTTAGATGACCCCTAAACATGAAAGGAATCATTAATAAGGATCTCAATATGT 3162
 Db |||||
 QY 1884 sProValLysLysIleLeuProSerLysAspAsn-ThrSerLysLeuAspThrAsn----- 1902
 Db |||||
 QY 3163 GGGCAGCTAAAGTTCTGATAGCTGAGAAACAATCTCAATCTTAAATTTAGGAAAAACAGA 3222
 Db |||||
 QY 1903 -----ValHisAsnAsnMetCysAspAsn-----IleTyrAsnMetLeuA 1916
 Db |||||
 QY 3223 CAAGAANAATAGAGCAGA----- 3242
 Db |||||
 QY 1916 snLysLysAspAspGlyLysGluAsnLysArgAspSerIleProAspTyrValSerVal 1936
 Db |||||
 QY 3243 --AATTGAATACACCACTCTGCTGCTCGCTGATCAAGCCATGATCAATTTGACATC 3300
 Db |||||
 QY 1936 laAspSerIlePheHisLeuAspAspCysLysAspThrGlnSerAsnAsnSerHisIleG 1956
 Db |||||
 QY 3301 AAGAAAAGTAAGAACCTGCTTCCACATTCAGGAGATGCTG----- 3341
 Db |||||
 QY 1956 lngluLeuGlnAspGluCysValGluLysIleAspMetIleAspLysAsnAspIleTyrT 1976
 Db |||||
 QY 3342 -----TTTGCAAGAAATCAATGCTGATGAGAGTACGATATTAACAATGAGTGC 3393
 Db |||||
 QY 1976 yrSerSerThrAsnLysLysAsnAspGluIleLysAsnArg-----AsnAspTyrP 1994
 Db |||||
 QY 3394 TCCATCACCCTTCTGAGCTCAAGGAAATCCAAAC-----CTAAATAATTTCTCAATTA 3450
 Db |||||
 QY 1994 helysAsnHisAsnIleCysAsnArgGlyValGlnAsnGlyAsnLysAsnLeuCysValA 2014
 Db |||||
 QY 3451 GCAGGAGATCTCTAAGAAATACATGTTTTCAGAACATCAACAAGACACACGCTGAA- 3509
 Db |||||
 QY 2014 snAsnLysAla-----AsnGluPheAsnAsnLeuHisLysLysGluAsnAsnI 2030
 Db |||||
 QY 3510 -----AACAGTGTCAATCAAGGAAGCGAACACATGTATCAAAACGACA 3552
 Db |||||
 QY 2030 leLeuAsnLysAspThrSerAlaIleAsnSerIleGlyAsnAsnLeuTyrLysAsnAsnA 2050
 Db |||||
 QY 3553 AGTAATGGAAACAAACACCTGAACAGACAGTCTCTAGTCAGAAATTTTCAACTACAA 3612
 Db |||||
 QY 2050 rg-----SerV 2052
 Db |||||
 QY 3613 GCAAAATTTGGCTTCACAGCAATTAGTCATGCACATAGA----- 3653
 Db |||||
 QY 2052 allysAsnIleAsnAspIleIleLysArgSerHisThrProGluIleValThrLysV 2072
 Db |||||
 QY 3654 -----AAGCTGACAAACAAAGCAGATACAAATGATTCATTTCTTGAGAGGAATGCAA 3708
 Db |||||
 QY 2072 alLeuLysIleGluGluThrIleAsnAspAsnThrLeuAsnValValHisLysLysGluT 2092
 Db |||||
 QY 3709 CACATCTCTAAAGAGAAAGAGGAGATATTAATCAATACCATTTAAACCGATAT 3768
 Db |||||
 QY 2092 hrHisLysGlnAsnLysGlyThr-Asp----- 2102
 Db |||||
 QY 3769 ATTCAATATGAAAGAGAAAGAGAAACAGAAATCATGAGAACACAGAGTAGAACTTC 3828
 Db |||||
 QY 2103 IleSerAsnGluPheLysGluAsnLysLysGluGlyLysSerAsnSerLeuLysArgTyr 2122
 Db |||||
 QY 3829 TTTGGAAACAACACACAGACTTTACTACACTCATCTAGAG-----GCCAGTCTAGCA 3885
 Db |||||
 QY 2123 LeuAsnLysThrCysAspAsnLeuAsnValThrSerArgIleAsnMetLeuAsnLysThr 2142
 Db |||||

QY 3886 TCACCTATGTTGAAATCTTACCAAGAGTCTGTG-----TCACAGAAATACTATTTAGAA 3939
 Db |||||
 QY 2143 SerProThrIleSerPheLeuAsnLysThrSerMetAsnAsnLysAsnValAspLysGlu 2162
 Db |||||
 QY 3940 GAAAT-----TCATGTTTCTTCTGAGCTTACAGCTACAGCATAAAATAAAGTGTGAAGATTA 3993
 Db |||||
 QY 2163 AspAsnTyrLysLysCys-SerAspThrIleLysGlyLysSerTyrAsnSerGluAspLe 2182
 Db |||||
 QY 3994 CTTGTTCCGAATTCATAGCTGCACAGATTCCTCATACCTGATGTCAGCAGACTCA 4053
 Db |||||
 QY 2182 uLeuTyrLys-----GlyAsnPheTyrLysAsnValArgAlaTyr 2196
 Db |||||
 QY 4054 TTCATC-----AACCAAGATCCG 4072
 Db |||||
 QY 2196 rLysIleGluAsnLysGlnPro 2203
 Db |||||
 RESULT 9
 Q34192
 ID Q34192 PRELIMINARY; PRT; 590 AA.
 AC Q34192;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE NADH dehydrogenase subunit 5.
 GN ND5.
 OS Crithidia oncopelti.
 OG Mitochondrion.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Crithidia.
 OX NCBI_TaxID=5657;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Dmitriy Maslov A.;
 RC STRAIN=S-068;
 RL Submitted (OCT-1990) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S-068;
 RA Maslov D.A., Horvath A., Gwang II K., Kolesnikov A.A.;
 RL Submitted (AUG-1991) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X56015; CRA39492.1; -.
 DR PIR; S34960; S34960.
 DR GO; GO:0005739; C:mitochondrion; IEA.
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0006120; F:mitochondrial electron transport, NADH to u. .; IEA.
 DR InterPro; IPR003916; NADHoxred5.
 DR InterPro; IPR001750; Oxidored_q1.
 DR Pfam; PF00361; oxidored_q1; 1.
 DR PRINTS; PR01434; NADHREDNASES.
 KW NAD; Oxidoreductase; Ubiquinone; Mitochondrion.
 SQ SEQUENCE 590 AA; 71650 MW; 1A2F3619B5B0C02E CRC64;
 Alignment Scores:
 Pred. No.: 3,82e-05 Length: 590
 Score: 146.00 Matches: 121
 Percent Similarity: 35.34% Conservative: 61
 Best Local Similarity: 23.50% Mismatches: 158
 Query Match: 2.03% Indels: 176
 DB: 8 Gaps: 27
 US-09-602-362E-22 (1-4115) x Q34192 (1-590)
 QY 3977 TTTTATGCTGAGCTTTCAGGAAGAAACATGAATTTTCTTCTAAATAGTATTCTGT--- 3921
 Db |||||
 QY 76 PheTyrCysAsnGlyPheTyrLeuPheIleLeuPheLeuLeuValPheCysPhe 95
 Db |||||
 QY 3920 -----GACACACATTTGTAAGATTTCAC 3894
 Db |||||
 QY 96 IleLeuPheTyrAlaPheTyrTyrMetTyrTyrAspLeuMetLeuLysArgPheAsn 115
 Db |||||
 QY 3893 ATAGGTGATGCTAGACTGGCTCTAGCATGAGTGTAGTAAAGTCTGCTGTGTTTCT 3834
 Db |||||
 QY 116 Ile-----PheTrpTyrPheValLeuCys 123
 Db |||||

QY 3833 CCAAAGAGTTTCTACTGCT-----TGTTCTCATGAGTTTCTG 3796
Db 124 MetAsnPhePheLeuLeuSerTyAspTyLeuThrAlaTyCysGlyTrpGluLeuLeu 143
QY 3795 TTTCTCTTTCTCTTTTTCATTAATGAAT-----ATACGGTTTTTAATAGTATTGTA 3745
Db 144 GlyLeuPheSerPhePheLeuLeuSerTyPheTrpTyArgPhePhe---AlaLeuLys 162
QY 3744 ATTAATATCTCCCTTTT-----TCTCTTTAGGAGATGTTGCATTTCTCTCAAGA 3691
Db 163 PheGlyPheLysSerPhePheLeuLeuSerLysLeuValLeuLeuLeuLeuSer--- 181
QY 3690 AATGAATATCATTTTATCTGCTT-----TGTTTGTGACGT 3655
Db 182 -----PheValMetThrPheLeuSerThrGlyTyGlyMetLeuAsnPhe 196
QY 3654 TTCTATGTGATGACTAATTTCTGTGAAGCCACAAATTTTGTGTAGTTGAATATTT 3595
Db 197 TyrPheValAsn-----Phe 201
QY 3594 CTGACTAGAGAC-----TCTGCTGTTTCAGGTGTTTGTTCATTAATCTTCTGCTTT 3544
Db 202 LeuCysValAspPheTyPheThrAlaPheMetLeuPheLeuLeuMetCysAlaPhe 221
QY 3543 GATACATGTGTGTCCTTC-----CTTCATGTACATGTTTTCACGTGTCTCTTTGTGATGT 3487
Db 222 ThrLysSerThrGlnPheGlyLeuHis----- 230
QY 3486 TCTGAACCAATGATTTCTTCTTAGACATCTCTGCTAATTGAGATAATTTTAGGTTT 3427
Db 231 -----IleTrpLeuProAspAla-----MetGluGly 239
QY 3426 GGATTCCTTTGAGCTCAGAAAGTGGTGAAGAGCATCATTTGTTAATATCGTACTCTCA 3367
Db 240 ProIleProValSerAlaLeuLeuHisAlaAlaThrLeuValValCysGlyLeuLeu 259
QY 3366 CATCAACATTCATTTCTTTGCAACAGCATCTCTCAATGCTGGAACAGGTCTTACTT 3307
Db 260 -ValSerPhePheTrpCysPheAspPheTrpLeu-----AlaTyPhe 274
QY 3306 TTTCTTGTATG-----TCACAAATGATCATGGCTTGACACGAGCAGTC----- 3261
Db 274 eTyProLeuIleGlyTrpSerSerLeuLeuValMetMetSerLeuCysValPheTy 294
QY 3260 -----AGATGTGTATTCATTTTCATTTTCGCTCTAG 3232
Db 294 rAsnPheAspAlaLysArgPheValAlaPheSerThrIleCysGlnIleSerPheSerMe 314
QY 3231 TTTTCTTTGCTGTTTTCCTCAATTTAGATGAGCATTTTCTCAGCTATCAGAACTT 3172
Db 314 tPheCysCysLeuCysLeuAspLeuTyValGlyCysLeuPhePheCysTyHisMetPh 334
QY 3171 TAGCTGCCACATATTTGATGATCCCTTTTGTAGTAAATGATTCCTTTTGTAGGTGATC 3112
Db 334 e-----TyLysAlaThrLeuPheLeuValLeuGly--Val 345
QY 3111 TAAGTTGAGCATCTTTCTTTTAAA----- 3088
Db 346 TrpIle-----HisLeuPhePheGlyLeuGlnAspValArgCysTyPhePheLeuTyPhe 364
QY 3087 -----ATCTTAAGTCCCTCAAGAATTTATTTTCTTTTCCCGATGTTGTTAGTG 3037
Db 365 CysGlyCysValLeuAlaArgMetLeuLeuValPheAlaLeuLeuAsnSerCysSerLeu 384
QY 3036 TGGCTATTTCCAGTTTGTGATGCCAATTCCTTTTCAACATGCAATTTTCATGTAAGATAA 2977
Db 385 TrpPheLeu-----CysGlyPheTy----- 391
QY 2976 TTTTTCATTTTCATGAGTTGAGAAACCTATTCAATTTACTTTCTACACCTTCAATTTCTATC 2917
Db 392 -----CysLysAspLeu-----LeuLeuCysThr----- 399

QY 2916 TTGTATCTGAGAGCCTTTCAAGTTGTGTCTTCACTTCTAACTTCTTATGCTGCTCTCGA 2857
Db 400 LeuMetLeuValSerPheHisPheLeuGluPhe---LeuPheValCysIle----- 416
QY 2856 TTTCTTAATTTCTTCTAAATTTTCTTAAATATATCGCATTTTCTCTCTCTCTCTTCT 2797
Db 417 -----PhePheLeuPhePheThrValIleTyAsnTyPheLeuLeuPhePheLeu 433
QY 2796 TGG-----TTTAAGTCAATCTACACTGACACTTCTTGTCCCATTTAACTTTGGTTCT 2743
Db 434 CysPheValPheLysCysPheCysLeuAlaAsp---CysLeuPheLeuLeuPheAspPhe 452
QY 2742 CTACTGTGATTTATTTCTTTTGTCTGACAGTCTTTTTCAGTACACAAACTCTTTTTC 2693
Db 453 GluCysCysLeuLeuTyCysPheLeuGlyLeuTyMetCysPheLeuLeuLeuPhePhe 472
QY 2692 ATTGTTCATTTTCTGTACGTGTTCACAGTGTCT 2648
Db 473 ValIleAsp--PheLeuTyIlePheValPheSer 483
RESULT 10
Q8IBJ6 PRELIMINARY; PRT; 418 AA.
ID Q8IBJ6
AC Q8IBJ6; (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein.
GN MAL7P1.142.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]_TaxID=36329;
RP SEQUENCE FROM N.A.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
Quail M., Barrell B.,
Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844506; CAD51004.1; -.
KW Hypothetical protein.
SQ SEQUENCE 418 AA; 51716 MW; 779413F255B58DEA CRC64;
Alignment Scores:
Pred. No.: 4,05e-05 Length: 418
Score: 145.50 Matches: 109
Percent Similarity: 38.69% Conservative: 62
Best Local Similarity: 24.66% Mismatches: 136
Query Match: 2.03% Indels: 135
DB: 5 Gaps: 24
US-09-602-362E-22 (1-4115) x Q8IBJ6 (1-418)
QY 3789 TTTCTCTTTTTCATATTGAAT---ATACGGTTTTTAAATGGTATTGTAATTAATCTCC 3733
Db 33 PheLeuTyPheLeuValHisPheLeuSerPhePheTyPhePheLeuValHisPheLeuSer 52
QY 3732 TCTTTTCTCTTTTAGGAGATGTTGATTTTCTCTCAAGAAATGAATATCATGTTAT 3673
Db 53 PhePheTyPheLeu-----ValHisPheLeuSer-----PhePheTyPheLeu 67
QY 3672 CTGCTTTTCTGTCAGCTTTCTATGTCATGACTAATTTGCTGTGAAGCCACAATTTTTC 3613
Db 68 ValHisPheLeuSerPhePheTyValHis-----HisSerTyPhe 81
QY 3612 TTGTAGTTGAATTAATTTCTGACTAGACTCTGCTGTTCAGGTGTTGTTGTTCCATTATCT 3553
Db 82 -----PheIle-LeuLeuArgLeuSerPhe-----IleLeuPheHisPhePhe 95
QY 3552 TGTC-----GTTTTCATACATGTTTCTCCTTTCATTTGACACTGTTT----- 3509
Db 95 eThrSerSerPheLeuPheHisPheThrSer-SerPheIleLeuPheHisSerP 115
QY 3508 -----TCAGTTCTCTTTTGTGATGTTCTGAACCAATGATTTCTCTTAGACATC 3457


```
Db 115 heThrSerSerPheIleLeuPheHisSerPheThrSerSerPheIleLeuPheHisSerP 135
QY 3456 TCCTGCTAATTGAGATAATTTTAGG---TTTGGATTTCCTTTGAGCTCAGAAAGTGGT 3400
Db 135 heThrSerSerPheIleLeuPheHisSerPheThrSerPheIleLeu-----151
QY 3399 GATGGAGCAGCTCATTTGTTAATATCGTACTCTCAGATCAACATTCATTTCTTCCAAACA 3340
Db 152 -----PheHisSerPheThrSerSerPheIleLeuPheHisSerP 165
QY 3339 GCATCTCCTCAATGTGGAGCAGGTTCTTACTTTTCTTGATGTCAACAATTGATCATGGC 3280
Db 165 heThrSerSer-----168
QY 3279 TTGTACAGCGAAGCCAGTCAGATGGTGTATTCAATTTCTGCCCTCTAGTTTCTTTGCT 3220
Db 169 -----PheIleL 171
QY 3219 GTTTTTCCTCAATTAGAAATGAGCAATGTTTCTCAGCTATCAGAACTTTAGCTGCCACCA 3160
Db 171 euPhe-----HisSerPheThrSerSerPheIleLeuPheHisSerP 185
QY 3159 TATTGAGATCCCTTTTAGTAATGATTCCTTTTCAGTTTATAGGTCATCTAAGTTTCAGCAT 3100
Db 185 heThrSerSerPheIleLeuPheHisSerPheThrPheIleIleHisPheIleSerPheP 205
QY 3099 CTTTTCCTTTAAATCTTAAGTCCTCAAGATTTATTTTCTTTTCTCGGATTTGGTTTGA 3040
Db 205 heTyVal-----HisSerTyTyPhe-----IleLeuLeuA 217
QY 3039 GTGTGGCTATCCAGTTTTCATGCGCAATTCCTTTTCAACATGCAATTTTCATGTAAAGAA 2980
Db 217 rgLeuSerPheIleLeuPheHis-----PhePheThrSerSerPhe-----I 231
QY 2979 TAATTTTTCATTCATGATGAGTTCAGAACTATTCAATTAATTTCTACCTTCACTTCAATCTT 2920
Db 231 leuLeuPheHisSer-----LeuThrLeuSerPheIleLeuPheHisSerL 246
QY 2919 ATCTGTATCTGAGAGCCTTTCAAGTTGTGTTTCACTTCTTAATCTTCTTCTATCTCTCTC 2861
Db 246 euThrLeuSerPheIleLeu-----PheHisSerLeuThrLeuSerPheIleLeuP 263
QY 2860 -----TCGATTCCTTCAATTCCTTCAATTTTCTTATATATATATCGG 2818
Db 263 heHisSerPheThrPheIleIleHisIleIleSerPhePheTyTyArgLeuPheT 283
QY 2817 CATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2758
Db 283 yrPheIleSer-----LeuLeuProA 290
QY 2757 TTAACCTTTGGTCTCTACTGTGATTTATTTCTTTTGTCTGCACAGTCTCTTTTTCAGTAC 2698
Db 290 rgSerPheTyTyPhe-IleLeuLeuArgSerSerPheIleLeu-----PheHisSerPhe 307
QY 2697 ACAAAACTCTTTTTCATTTCTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2647
Db 308 ThrPheIleIleHisIleIleSerPhePheTyValCys-HisSerLeuAsnPheIleLe 327
QY 2646 TTGAAGTTCCCTTGCTTTCACAGATGAATGATGACCAAGATTTTGATAGGCTAGTGAA 2587
Db 327 uLeuArgLeuSerPheIle-----LeuPheHisSerPheThrPheIleValHi 343
QY 2586 TCTTCTAAATTTCCACTATTATTTCTCTTTCTTTTGTGTGTAGCTGGGTACACACCA 2529
Db 343 sPheIleSerPheTyTyPheTyArgSerTyTyPhe-----PheThrPro 358
```

RESULT 11

```
ID Q86KF8
AC Q86KF8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
```

DE Similar to Dictyostelium discoideum (Slime mold). Interaprin.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Eukaryota; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=22092622; PubMed=12097910;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
Lehmann R., Baumann C., Parra G., April J.F., Guigo R., Kumpf K.,
Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum";
RL Nature 418:79-85(2002).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX Baumann C.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC115680; AAO51092.1; -.
SQ SEQUENCE 1781 AA; 209555 MW; 6483EE19ED1B76A8 CRC64;

Alignment Scores:

| | | | |
|------------------------|----------|---------------|------|
| Pred. No.: | 7-25e-05 | Length: | 1781 |
| Score: | 144.00 | Matches: | 237 |
| Percent Similarity: | 38.21% | Conservative: | 202 |
| Best Local Similarity: | 20.63% | Mismatches: | 395 |
| Query Match: | 2.03% | Indels: | 316 |
| DB: | 5 | Gaps: | 49 |

US-09-602-362E-22 (1-4115) x Q86KF8 (1-1781)

```
QY 817 AGCTGTGGTGGAAAAACACCTGATGAGCTGCATCTTGGT-----GGA 858
Db 272 LysLeuLysGluSerGluIleLeuMetAspLeuGlnIleGlnValSerAspLeuGln 291
QY 859 GAAACATCTGCAAAATTCATTTGCGAGAAAGGACATCTCGAAAGTTGGAAGTCAGCAG 918
Db 292 SerThrSerAspAspAsn-----GluArgTyGlnSerLeuIleSerGluTyGln 309
QY 919 AAAACACCTAGGAATTCAGATCTGCAAAAACATCTGAGAAATTTACGGCCAGCA 978
Db 310 GluAsnLeuLysSerValSerGlnLeuAsnLysGluIle-----322
QY 979 AAGGAAGACCTAGAGATCGCAGAGAGAGAGAGACACACCTAGGAAATATGAGTCC 1038
Db 323 AsnGluLysLeuLysAsnGluArgThrGluSerLysSerLeuIleAspLys-IleSerLy 342
QY 1039 CGAAAAGAAACACTGAGAAATTCAGTGGCAGAGAAAGGAA-----GACCTAGG 1086
Db 342 sGluAsnAsnAspTyTyGlnIleLeuIleThrGluGlnAspLysAspLeuLys 362
QY 1087 AAGACGCATGGGAGAAAA-----GAAACCTGTAAAGATGATGCGTGC 1131
Db 362 sSerGlnLeuAspThrLysSerAsnAsnTyTySerLysLeuLeuAspLeuSerLeuSe 382
QY 1132 AAGAGTAACATCTATAAATCTTAAAGTTTGAAGAAAGAAATCTAAGATGTTGTCATGCCA 1191
Db 382 rLysLeuAsnTyTyAsnLysLeu-AsnLeuGluPheAsnAsnIle-----396
QY 1192 CAAAAGATCATCTACAAAGCAGTGCATGATCAGAGTCCCATCATGATCCAAACAAGG 1251
Db 397 -----SerAsnGluTyTyGlnIleGluLysGlnLeuSerPheAspPheAsnGluL 414
QY 1252 GAAGATCAAGAATATCTTTGATTTCTCGGCTCTTTTGAAGTTCTGCAAGATTCAGTGG 1311
Db 414 euLysLysLysLeu-----AspAspGlnIleGluArgAsnSerAsnGlnSerLeu- 431
QY 1312 TATACCTGAGTCTATATATAAAGTAATGAGATAAATAAGAGTAGAAGACCTCTCTAAA 1371
Db 432 -----LeuThrGlnArgIleAsnGlnIleIleAspLysGlnLysGluIleAspA 448
QY 1372 AGCCATCTGCTTCAAGCCGCCCATTTGAACCAAAATCTTTTCCAAATATAGCCTTGAATG 1431
Db 1431
```

Df

448 snleuasnSerAspleuGlnSerleuAasnAsplys-----GlnSerAspSerAenA 465

1432 AAGAATGAACAAACATTGAAGCAGATCGTGTTCCACCAGAAATCCAACAACAAAGCAC---- 1487

Df

465 spGlnileAsnThrileThraAsnGluasnaAsnAspleuLysileLysLeuGluasnerI 485

1488 -----TATAAGAAAATTCCTGGATTCTGGAGTTCAGAGTCTCTGGAGACTG 1527

Df

485 leuleuleSerAsnHisGlnAsplysleuthrThrAsnLysLysGluLeuGluSerL 505

1528 TTTACAGAGAGTTCGTGTTTACCAGGCCTACATCAA-----AAAG 1569

Df

505 eutySerilleleGluAenLeuAasnSerThrHisGlnAspGlnLeuLysGluLysGluA 525

1570 AATAGATAAAAAATGGAAATAGAA-----GAGTCCCTTAATAAGTGGTCTTCAGAG 1623

Df

525 snGlnlelGuInMetLysleuAspGlnSerGluSerleuAasnLysPheGlnGlu-Lys 544

1624 GCTACCTCGGAATGAAATTTCTATTCACCTAAAGCCTAGAAATGAAGACATGCCAACATTT 1683

Df

545 ilethrThrLeu-----AsnHisThrIleValAsnIle 555

1684 CAAAGCGAGCCTCCGGGAAGCATCTGCT- - - - -CGAGCCTCCACCTCAATCAAAAGT 1737

Df

556 AsnSerGluLysAspSerLeuAasnLysleuIleAsnAspTy-SerGlnGlnleAsnGlu 575

1738 CTCTCCCAATAAAGCCTTGGAAATGAAAAATCAAAAAACATGGAGCGAGATGAGAACTCCCA 1797

Df

576 LeuAsnLysileAsn-----AsnGlnLysAsn-Gln----- 585

1798 TCAGATCCAAACAAGGACTATAAGAAAATTCGGGACTACTGGAGTCTCTGTAGACTGT 1857

Df

586 -----GluAenGlnGlnLeuLeuSerSerIle----- 594

1858 TTCACAGAGAGTTCGTGTTTACCAGGCTCGCTCAAAAAGAAATGATAAAATATGAGAA 1917

Df

595 ---ArgGluAspCys----- 598

1918 AATTAGAAGGTCCTCGTTTAAAGTGGTCTTCTGAGGCTAACTGGGAATGAAAGTTCTATT 1977

Df

599 -----GlnThrIleAs 602

1978 CMACTAAAGCCTAGAAITGTATGATGACATGQA--ACTTCAAAGCAGGCTCCGAGAGCC 2034

Df

602 nGlnLeuLysSerThrIleLysSerLeuGlnAspGluSerAsnSerLeuSerThrGluIl 622

2035 ATCTGCTCTCGAG-----CCTCCATTGCAATCAAAAGTCTGTCCAAATAAACCTTG 2085

Df

622 eGluAlatleLysLeuGlnLeuAenGlnLeuSerThrIleThrIleProGluLys----- 640

2086 GAATTAGAATGAACAACAACTTGGCAGATGAGACTCCATCAATCAAAACAAAGGA 2145

Df

641 -AspGlnGluLeuSerAenLysGluA-zgThrIleGlnGluPheGlnValLysThrGlnGl 660

2146 CTATGAGAAAGTT-----CTTGGATTCTGGAGTCTCTGTAGACTGTTC 2190

Df

660 nLeuLysGlnThrIleGlnGlnAenGlnLeuThrIleAenGlnHisLeuThrThrIleAs 680

2191 CAGAAGGATGCTTTTACCAGGCTACATCAATCAAAAGAAATAGATAAAATATGGAAAT 2250

Df

680 pAsnGlnSerValAspIleAenSerLeuAenGluLys----- 692

2251 AGAAGAGTCTCTGATATAATGAGTTTTCTGAAGCTCCCTGAGAANTGAAAGTTCTATTCCA 2310

Df

693 -----LeuValGlnLeuAenAspGluSerIleLysLysGlnGlnSerIleHi 708

2311 CTAAGCCTTGAATGTATGGCATCAACCTTCAAGACAGGCTCC----- 2356

Df

708 s---SerLeuSerLeuGlnVallleGluLeuAenLysLysLeuSerGluLysAspAspGI 727

2356 ----- 2356

Df

727 ntYrAsnGlnSerLeuGluSerIleAspGlnLeuThrSerGluLeuGlnLeuLysGlnAs 747

```

QY 3282 CATGATCAATTGTGACATCAAGAAAAAGTAAGAACCTGCTTCCACATTGAGGAGATGCTG 3341
Db 1060 -----GlnAspLeuSerThrSerAspTyrLysIleGlnGlnLeuGln 1073
QY 3342 TTTCAGAAAGAAATGAATGTTGATGTCAGAGTAGCATATTAACAATCAGTGTCCATCAC 3401
Db 1074 IleAspLeuGlnIleAspLysAspGluIleIleLysLeuGluThrlle----- 1090
QY 3402 CACTTTCTGAGCTCAAGGAAATCCAAACCTTAAATTAATCTCAATTACAGAGATGT 3461
Db 1091 -----SerGlnArgAsnGlnSerIleLys-----GluSer 1100
QY 3462 CTAAGAGAAATATGTTTTCAGAACATCACAAAGAGACAAACGCTGAAACACGTGCAAT 3521
Db 1101 LeuValLysCysAsnAspLeuGlnAspGluThrSerLysLeuAsnAspAsnLeuGln 1120
QY 3522 GAAGGAGCAACACATGTTATCAAAACGACACAGATATGAAACAAACACCTGACACGAC 3581
Db 1121 LeuAsnSerThrIleThrAspTyrGlnSerGlnIleThrGluSerAsnGluAsnValGln 1140
QY 3582 AGTCTCTAGTCAGAAATATTTCAACTACAAAGCAAAATTTGTGCTTCACAGCAATTAGT 3641
Db 1141 SerLeuGlnAsnGlu-LysAsnGlnLeuGlnLeuGluLeu-----AspGlnLeuLys 1157
QY 3642 CATGCAATAGAAGCTGACAAACAAAGCAGATACAAATGATATTCATTTCTTGAGAGGA 3701
Db 1157 sginArgile---SerGluGlnHisAspAspIleThrLeu----- 1169
QY 3702 AATGCAACATCTCTCTTAAAGAGAAAGAGAGAGATATTAATTAACATACCAATTAAAA 3761
Db 1170 -LeuAsnSerIleGluPheGluLeuAsnArgLysIleSerAsnTyrGlnSerAspIleLys 1189
QY 3762 ACCGTAT-----ATTCAATATGAAAGAGAGAAA-----GAGAA 3794
Db 1189 sGluTyrAspAsnAsnIleLysValIleGlnAsnGluLysAsnGlnLeuGluLe 1209
QY 3795 ACAGAAACTCATGACACAGACGAGTAGAACTTCTTTGGA---GAAACACACAGACAGAC 3851
Db 1209 uAspGlnLeuLysGlnValLeuSerAspLysGlnAspGlyValSerThrLeuAsnSerTh 1229
QY 3852 TTCTACTCACACTC 3864
Db 1229 rLeuLeuGluLeu 1233

RESULT 12
Q81BY8 PRELIMINARY; PRT; 2910 AA.
AC Q81BY8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN PF07_0042.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=36329;
RN [1]_SEQUENCE FROM N.A.
RP Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844506; CAD50853.1; -
KW Hypothetical protein.
SQ SEQUENCE 2910 AA; 350698 MW; B9F142C5A895C8C1 CRC64;

```

Alignment Scores:

```

Pred. No.: 7,86e-05 Length: 2910
Score: 144.00 Matches: 223
Percent Similarity: 38.07% Conservative: 200
Best Local Similarity: 20.07% Mismatches: 357
Query Match: 331 Indels: 51
DB: 5 Gaps:

```

```

US-09-602-362E-22 (1-4115) x Q81BY8 (1-2910)
QY 1133 AGAGTAACATCTAATAAAGTTTGAAGAAAGAAATCTAAGATGTTGCATGTCAC 1192
Db 1072 LysIleLysTyrAsnAspIleLysThrAsnLeuGluA:ggLulYsIleAsnMethis 1091
QY 1193 AAAAGAATCATCTACAAAGCAAGTCCCATGATCAGAGGTCCCATCAGATCCAAACAGGG 1252
Db 1092 MetAsnIleGluLeuGluLysGluArgLysIleAsnAspGluMetMetIleLysPheGlu 1111
QY 1253 AAGATGAAGAATATTCTTTGATTCCTCGGTCTCTTTGAAGTCTCTCAAGATCAAGTGT 1312
Db 1112 Lys---LysAsnLysMet---TyrAspAspIleAsnLysGluLeuGluAsnLysLys 1129
QY 1313 ATACCTGAGTCTPATATATAAAAGTAATAGATAAATAAGAGTAGAAGCTCTCTATAAA 1372
Db 1130 GlnAsnLysGluIleTyrMet-----GluLeuGluAsn---IleLys 1142
QY 1373 GCATC-----TGCCTTCAAGCCCATTCGAAGCAAACTCTTTCCAAATAAGCC--- 1423
Db 1143 AsnIleAsnSerAspMetAsnPheAlaLeuGluAspLysSerValGlnTyrAspGlnLys 1162
QY 1424 -----TTTGAATGAAGAATGAACAAACATTCGAAGCAGATCCGTGTTCC----- 1466
Db 1163 CysIleAspPheGlu---LysLeuAsnLys-MetTyrGluAspThrCysLysGluLeuAs 1181
QY 1467 -----CACGAATCCAAACAAAGGACTATTAAG----- 1493
Db 1181 nAsnLysAsnValLeuTyrGluGlnMetAsnThrGlnLeuTyrLysGluLysIleAsnAs 1201
QY 1494 ---AAAATTTCTGGATTTCTAGAGTCTCTCGAGACTGTTTACAGAGGATGTTGTTTAC 1549
Db 1201 nAspGluIleCysThrGluLeuGluLysIleLysAsnIleAsnLysAsnIleGlnIleAs 1221
QY 1550 C-----AAGGCTACATCAATAAAGATAAGATAAATAAATAAGATAAGATAAGATA 1597
Db 1221 nLeuGluAsnGluLysThrAsnGluGlnLys-AspLysGlnIleGluLysGlnLys 1240
QY 1598 GTCCCTATAAAGTGTCTTCTGAAGGCTACCTCGG-----AATGAATTT 1642
Db 1241 MetAsnLysGluMetValIleGlnLeuGluLysGluLysIleIleAsnAsnGluIle 1260
QY 1643 TCTATTCACATAAAGCCCTAGAAATTTGAAGACATGCAAACTTTCAAAGCGAGCTCCGGAA 1702
Db 1261 IleThrGlnLeuGlu-----AsnGluLysLeuAsnAsn 1271
QY 1703 GCCATCTGCTTCGAGCCTCCACTGAAATCAAAAGTGTCTCCCAATAGCTTGAATGA 1762
Db 1272 LysIleIleTyrGluAsp-----IleGluLys 1280
QY 1763 AAAATGAAA---AACATGGAGGAGATGAGAACTCCCATCAGAAATCCAAACAAAGGACTA 1819
Db 1281 LysAsnLysLeuAsnSerGluLeuGluAsn---TyrGluAsnGlnGlnLysLysIle 1299
QY 1820 TAAGAAAATTTCTGGGATCTGAGTCTCTGTAGACTGTTTTCAGAGAGGATGTGTTT-- 1877
Db 1300 AsnGluMetIleIleGlnLeuGluLysGlu-LysIleIleAsnAsnGluIleIleGln 1319
QY 1878 -----ACCAAGCTCGCTCAAAAGAGATAGATAAATAATGGAATAATGGAAGG 1927
Db 1319 nLeuGluAsnGluLysGlySerValLysLys-----IleAsnThrGluLeuGlu-- 1335
QY 1928 GTCCCTGTTAAAGTGTCTTCTGAG-----GCTAACTGGGAATGAAAGTTCTA 1975
Db 1336 -AsnIleLysLysMetAsnAspGluMetAsnGluLysLeuAsnLysGluGlnLysIleAs 1355
QY 1976 TTCCAATAAGCTAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2035
Db 1355 nAsnAspLeuGlnLeuAspLeuGlu----- 1363
QY 2036 TCTGCCTTCGAGCCTCCATTTGAATCAAAAGTCTGTCCAAATAAATCACTTGA-- 2087
Db 1364 -----AsnGluLysAsnGluLysAspHisIleAsnLysGluPh 1376

```

QY 2088 -----ATTAGATGAACAAACATTTGAGGACGATGAGAACTCCCATCAATC 2134
Db eGluLysGluLysIleLysAsnIleGlnLeuLysLysAsnMetAspGluGluLysLys 1396
QY 2135 CAAACAAAGGACATATGAGAAAGTTCTGGGATTTCTGGAGTCTCTGTAGACTGTTCCAGA 2194
Db : : : : : 1400
QY 2195 AGATGGTGTATACCAGGCTACACATCAAAAGAAATAGATAAATAATGAGAAATAGAA 2254
Db : : : : : 1401
QY 2255 GAGTCTCTGATATGAGGTTTCTGAGGCTCCCTGAGAAATCAAAAGTTCTATCCACTAA 2314
Db : : : : : 1412
QY 2315 AGCCTTGATGATGGCATGMACTTTCAAGAGGCTCCCGAGAGCATCTGCTTCG 2374
Db : : : : : 1425
QY 2375 AGCCTGCATGAAATGAAAGTCTGTTCCAAATAAACCTTGGAATTAAGAAATGAAACAA 2434
Db : : : : : 1443
QY 2435 TTGAGACGATCAGATTTCCCTTCAGAA---TCAAAACAAGAGGTTGAGAAATCTG 2491
Db : : : : : 1456
QY 2492 CGAT : : : : : 2533
Db : : : : : 1474
QY 2534 TGATCCAGGCTACACAAAGAAAGGATAAATAGTGAATAATGAGAAATTCAC 2593
Db : : : : : 1492
QY 2594 GCCTATCAAAATCTGGTACAGTTTCATCTTGAAAGGCAAGGAAGTTCAAAAGACAC 2653
Db : : : : : 1498
QY 2654 TGTAACACGTACAGGAAATGGACAAATGAAAGAGTTTGTGTACTGAAAGAACT 2713
Db : : : : : 1516
QY 2714 GTCAACAAAGAAATAAATACAGTAGAGAACCAAAAGTTAAATGGGAACAAGAGTCT 2773
Db : : : : : 1536
QY 2774 CGAGTGTAGATTGATTAACCAAGAAAGAAAGAGAGAAATGCCGATATATTAAGAA 2833
Db : : : : : 1554
QY 2834 AAAATTAGAGAAATTAAGAATCGAGAGCAGCATAGAAAGAGTTAGAAATGAAACACA 2893
Db : : : : : 1572
QY 2894 ACTT-----GAAAGCTCTCAGATACAGATTAAGATTAAGAGTTGAGAAAT 2944
Db : : : : : 1592
QY 2945 TTGAATAGTTTCTCACTCATGAAATG----- 2972
Db : : : : : 1612
QY 1612 eAspThrSerAsnValAsnGlnMetAsnAspGluHisValAspGlnMetAs 1632
QY 2973 -----AAAAT 2977
Db nAspAlaGluSerGluAspAsnThrPheLeuGluLeuGlnLeuGluLysValysGlnVa 1652
QY 2978 TATTCATCATGAAATTCATGTTGAAAGAAAGAAATGCCATGCAAAAGCTGGAATGCCAC 3037
Db : : : : : 1652
QY 1652 lAsnIleAspMetIleIleGlnLeuLysLysAsp-----LysLysArgIleAspG 1669

QY 3038 ACTAAACACCAATCCAGAAAGAAATAAA-----TCTTTGAGACTT 3082
Db : : : : : 1669
QY 3083 AAGATTTTAAAGAAAGATGCTGAACCTTAGATGACCTCT---AAAAGTGAAGAAATCAT 3139
Db : : : : : 1699
QY 3140 TACTAAAGGGATCTCAATATGTGGCAGCTAAAGTTCTGATAGCTGAGAAACATGCTC 3199
Db : : : : : 1708
QY 3200 ATTCTAAATTTAGGAAAGAAACAGACAAAGAAACTAGAGGAGAAATTTGAATACCATCC 3259
Db : : : : : 1713
QY 3260 TGACTGGCTTCGTGTACAGCCATGATCAATGTGACATCAAGAAAGATGAACACCTG 3319
Db : : : : : 1733
QY 3320 CTTCCACATTTGAGGAGATGCTGTTTTCAGAGAAATGAATGTTGTGAGAGTACGATA 3379
Db : : : : : 1753
QY 3380 TTAACAATGAGTCTCCATCACCCTTTCTGAGCTCAAGAGAAATCCAAACCTA----- 3434
Db : : : : : 1772
QY 3435 -----AAAATTTCTCAATTAGCAGGAGATGCTA-----AGAGAAAT 3472
Db : : : : : 1787
QY 3473 ACATGTTTTCAGACATCAAAAGAGACAACTGAAACAGTGTCAATGAGAGGACGA 3532
Db : : : : : 1806
QY 3533 ACATGTATCAAAACCAAGATAATGGAACAA----- 3566
Db : : : : : 1826
QY 3567 -----ACACCTGAACAGCAGAGTCTTAGTCAGA 3595
Db : : : : : 1846
QY 3596 AATTATTTCAACACAGCAAAATTTGTGCTTTCAGACAAATTAGTCACATAGAAA 3655
Db : : : : : 1866
QY 3656 GCTCACAAACAGACAGATAACATGATTTCTTTCAGAGGAAATGCAA----- 3708
Db : : : : : 1886
QY 3709 -----CACATCT 3715
Db : : : : : 1905
QY 3716 CCTAAACAGAAAGAGAGATTAATTAATACATCACTT----- 3756
Db : : : : : 1925
QY 3757 -----TAAACACCGTATATTCAATATGAAAGAGAAAGAG----- 3792
Db : : : : : 1945
QY 1945 ysIleAsnGluGluGlnTyrIleGlnLeuLysAspLysGluIleAsnSerMetV 1965
QY 3793 -----AAACAGAACTCATGAGA 3811
Db : : : : : 1985
QY 3812 CAAGCAGTAGAAATCTTTTGGAGAAACACACAGACACTTTTACTACACTGCTAG 3871
Db : : : : : 1985
QY 3872 AGGCCACTGAGTACCTATGTTGAAATCTTACCAAGTCTGTGTACAGAACTACTA 3931

Db 1999 luyGln---AlaAsnLysLeuAsnLeu-----LeuAsnGlnAsnLysL 2015
 QY 3932 TTTTAGAGAA 3942
 Db 2015 ysileAsnGlu 2018
 RESULT 13
 Q8CIP5
 ID Q8CIP5 PRELIMINARY; PRT; 1345 AA.
 AC Q8CIP5
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Dispatched B.
 GN DisP2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=129;
 RX MEDLINE=22259598; PubMed=12372301;
 RA Ma Y., Erkner A., Gong R., Rao S., Taipale J., Basler K., Beachy P.,
 RT "Hedgehog-Mediated Patterning of the Mammalian Embryo Requires
 RT Transporter-Like Function of Dispatched.";
 RL Cell 111:63-75(2002).
 DR EMBL; AV150699; AAN52162.1; -.
 DR MBL; MGI:2388733; Disp2
 DR InterPro; IPR000731; SSD_5TM.
 DR PROSITE; PS50156; SSD; 1.
 SQ SEQUENCE 1345 AA; 147943 MW; 931738C26FD3A57F CRC64;

Alignment Scores:

Pred. No.: 0.000173 Length: 1345
 Score: 140.00 Matches: 251
 Percent Similarity: 31.90% Conservative: 171
 Best Local Similarity: 18.97% Mismatches: 414
 Query Match: 1.95% Indels: 489
 DB: 11 Gaps: 60

US-09-602-362E-22 (1-4115) x Q8CIP5 (1-1345)

QY 3355 ATTTCCTTTGAAACAGCATCTCTCAATGTGGAAGCAGGTCTTACTTTTCTTGATGT 3296
 Db 245 ValPheValSerThrSerGlySerLeuTrpAsnLeuGlnAlaIleHisSerMetCys 264
 QY 3295 CACAATTGATCATGGCTTCTCAGGAGCAGCCAGTCAGGATGGTGATTCAATTTCTGCCT 3236
 Db 265 Arg-----IleGluGlnGluIleArgSerHisIleSerPheGlyAla 279
 QY 3235 CTAGTTTCTTGTCTGTTTTCTCAATTTAGAATGACATTGT- 3191
 Db 280 Leu-----CysGlnArgSerAlaAlaAsnGluCysCysProSerTrpSerLeu 295
 QY 3191 ----- 3191
 Db 296 GlyAsnTyrLeuAlaValLeuSerAsnArgSerSerCysGlnAspThrGlnAlaAsp 315
 QY 3190 -----TTCTCAGCTATCAGAACTTTAGCTGCCACATATTGAGATCCCTTTTA 3143
 Db 316 ThrAspArgThrLeuAlaLeuLeuArgPheCysAlaThrPheTyrHisArgGlyValLeu 335
 QY 3142 GTA-----ATGATTCCT--- 3131
 Db 336 ValProAlaCysValGlySerSerGlnAspLysProProPheCysAlaGlnValProAla 355
 QY 3130 -----TTTCAGTTTTAGGGTCAATCAAGTTTCAGCATCTTTT 3095
 Db 356 LysCysThrGlySerAsnValValTyrGluPheLeuHisTyrLeuLeuAspArgAspPhe 375
 QY 3094 CTTTAAATCTTAAGTCTCTCAAGATTTATTTCTCTTCTCGGATTTGTTAGTGTG 3035

Db 376 -----LeuSerProGlnThrAlaAspTyrGlnValProSerLeuLysPheAlaLeu 392
 QY 3034 GCT---ATTCCAGTTTTCATGGCAATTCCTTTTCAACATGCATCAATTCATGTAAGAATA 2978
 Db 393 LeuPheLeuProIleIleLysThrSerSerLeuLeuAsp----- 405
 QY 2977 ATTTTCATTTTCATGAGTTGAGAAACCTATTCAAAATTACT---TTCTACACCTTCAATTCCT 2921
 Db 406 IleTyrLeuAspGlyLeuGlyAspProIleLysValSerAspAsnTyrThrSerIleSer 425
 QY 2920 -----TATCTTGATCTGAGAGCCTT 2900
 Db 426 GlyMetAspLeuGlyLeuLysProArgLeuLeuLysTyrTyrLeuAlaGluAspThrMet 445
 QY 2899 TCAAGTTGTTTTCATCTTAACCTCTTCTATGCTCTCGATTTCTTCTAATCTTCTCT 2840
 Db 446 TyrProLeuIleAlaLeuValIle----- 454
 QY 2839 AATTTTCTTTTAATATATATCGCATTTCTCTCTCTCTCTTCTTCTGTTTAAAGTCAATCTA 2780
 Db 455 -----PhePheGlyMetSerLeuTyrLeuArgSerLeuPhe----- 466
 QY 2779 CACTGCAGACTCTTTGCCATTTAACTTTTGGTTCTCTACTGTGATTTATTTCTTTGT 2720
 Db 467 -----Ile-ThrPheMetSerLeuLeuGlyValLeuGlySerLe 479
 QY 2719 TCTGACAGTCTTTTTCAGTACACAAAACCTCTTTTTCATTTGTTCCATTTCTCT---GTAC 2663
 Db 479 uMetValAlaTyrPheLeuTyr-HisValAlaPheArgMetAlaTyrPheProPheVala 499
 QY 2662 GTGTTTCACAGT---GTCTTTTGAAGTTCCTCTTCCCTTTCACAGATGAATTCACCAAG 2606
 Db 499 snLeuAlaAlaLeuLeuLeuSerGlyValCysValAsnTyrThrLeuIlePheLeuAla 519
 QY 2605 ATTTTGTAGGCTA-----GTGAATC 2585
 Db 519 spMetTrpArgLeuSerArgGlyGlnValProSerGlyGlyMetProHisArgValGlyA 539
 QY 2584 TTCTAATTTTCCACTATTTCCTTTCTTTTGTGTAGCTGGGTACACACCATCCT 2525
 Db 539 rgThrMetHisHisPheGlyTyrLeu-LeuLeuValSerGlyLeu----- 553
 QY 2524 TCTGTAAACAGTCCACGAGACTTCAGAAATCCAGAAATTTCTCAACCTTCTT--- 2471
 Db 554 -----ThrThrSerAlaAlaPheTyrGlySerTyrLeuSerArgLeuProAla 569
 QY 2470 -----TGTTTTGATTTCTGAAGGAATCTGATCTGTCTCAATGTTGTTTCAATTTAAT 2417
 Db 570 ValArgCysPhe----- 573
 QY 2416 CCAAGTTTATTTTGGAACAGACTTTTTCATTTCAATGCAGGCTCGAAGCAGATGGCTCG 2357
 Db 574 AlaLeuPheMetGlyThrAlaValLeuValHisMetGlyLeuThrLeuLeuTrpLeuPro 593
 QY 2356 GGAGGCTGCTTTGAAAGTTTGATGTCATCCATCAANTCAAGCTTTAGTGAATAGACTTT 2297
 Db 594 AlaThrValVal-----LeuHisGlu-----ArgTyr 602
 QY 2296 CATTTCTCAGGAGCGCTTCAGAAACCTCATTTATCAGAGACTCTTCTATTTTCCATTTT 2237
 Db 603 LeuAlaHisGlyCysValAlaGlnAlaHisGlyGlnArgGlyGlySeraspProLeuArg 622
 QY 2236 TATCTATTCTTTTGTAGTGTAGCTGGGTAACACCATCTCTTCTCGAAACAGTCTACAG 2177
 Db 623 LeuLeuAlaLeu-----His-----Arg 629
 QY 2176 AGACTCCAGAATCCCAAGAACTTTCTCATAGTCCTTTGTTTGGATTGATGGAGTTCTTC 2117
 Db 630 ArgIleArgIlePheArgLysIleIleSerIleLeuSerArgLeuLeuPheGlnArgLeu 649
 QY 2116 ATCTGCTCAATGTTTGTCTTCAATTTCTTAATTCCAAGGTTTATTTGGACAGACTTTTGA 2057

Db 650 LeuProCysGlyValIleIysPhe-----ArgTyrIleThrIleCysThrPheAla 666
QY 2056 CAATGGAGCTCGAAGGAGATGGCTTCGGAGGCTGCTTGAAGTTTGCATGCAATC 1997
Db 667 AlaLeuAlaAlaGlyGlyAlaTyrIleGlyGly-----Val 678
QY 1996 AATTCTAGGCTTTAGTTG----- 1979
Db 679 SerProArgLeuGlnLeuProIleLeuLeuProIleLeuGlyGlnPhePheArgSerSer 698
QY 1978 -----GAATAGAACTTTCATTTCCAGATTAGCTCAGAGAACCACTTT 1937
Db 699 HisProPheGluArgPheAspAlaGluTyrArgGlnGlnPheLeuPheGluAspLeuPro 718
QY 1936 AACAGGACCCCTCTCAATTTTCCATTATTTTATCTATCTTTTGTG----- 1892
Db 719 ProAsnGluGlyGlyAsnLeuProValValLeuValTyrGlyIleLeuProValAspThr 738
QY 1891 -----AGCGAGCCTTGGTAAACACAACTCTCTGTGAAC 1856
Db 739 SerAspProLeuAspProArgThrAsnSerSerValValSer--AspProAspPhe---- 756
QY 1855 AGTCTACAGACACCCAGATCCAGATTTCTTATAGTCCCTTGTTCGATCTGATG 1796
Db 757 --SerProSerSerProGluAlaGlnGluTyrLeuLeuAlaLeuCysHisGlyAlaGlnA 776
QY 1795 GGAGTTCTCATCTGCTCCATGTTTTCATTTTTCATTCACAGGCTTATT---TGGACA 1739
Db 776 snGln-----SerPhePheGlyAspGlnProGluGlyTyrProT 789
QY 1738 GACTTTTCATTTTCAGTCGAGACTCGAAGCAGATGGCTTCCGGAGGCTGCTTGAAGT 1679
Db 789 hrLeuCysLeu---ValGluAlaLeuGlnGlnTyrMetGluSerProSerCysGlyArgL 808
QY 1678 TTGCA----- 1674
Db 808 euGlyProAspLeuCysCysGlyGlnSerGluPheProTyrAlaProGlnLeuTyrLeuH 828
QY 1673 --TGTCTTCAATTTAGCTTTAGT-----GGAATAG 1644
Db 828 isCysLeuLeuMetAlaLeuGluGlnSerProAspGlyThrArgAspLeuGlyLeuA 848
QY 1643 AAATTTTCATTCGAGGTAGCTTCAGAGACCACTTTTATAGGACTCTCTATTTTCCA 1584
Db 848 rgPheAspThrHisGlyAsnLeuAlaLeu-----ValLeuLysPheGlnT 864
QY 1583 TTTTTCATCTATCTTTTCATGTCGAGCTTGGTAAACACAACTCTCTGTAAACAGT 1524
Db 864 hrAsnLeuProTyrSerThrGluTyrGlyPro-----ValHisPheTyrThrGluI 882
QY 1523 CTCCAGAGACTCTAGAAATCCCAAGAAATTTCTTATAGTCTTGTGTTGGATTCTGGTGGGA 1464
Db 882 leSerArgTyrLeuSerThrGluMetSerLysAlaProProGlyLeuAsnGlnGlyTyrP 902
QY 1463 ACAGGATCTCTCAATGTTGTTTCATTTTCATTCAGAGCTTATTTGGAAGAGTTT 1404
Db 902 heThrSerAsnGluLeuTyrSerLeuGlnHisSerLeuSerThrGluProAlaVal 922
QY 1403 TGCTTTCAATCGCGCTTGAAGGAGATGGCTTTTAGGAGGCTTCTACTCTTATTTAT 1344
Db 922 alLeuGlyLeuAlaLeu-----AlaLeuAlaPhe-----AlaThrLeuLeuLeuS 937
QY 1343 CTCATTACTTTTATATAGACTCAGGTATACCACTTGAATCTTTCAGAACTTCAAGA 1284
Db 937 erThrTyr-----AsnValProLeuSer----- 944
QY 1283 GACCGAGAAATCAAGAAATATCTTCATCTTCCCTTCTTTTGGATCTGATGGACCTCGA 1224
Db 945 -----LeuPheSerValAlaAlaValAlaGlyThrVal 956
QY 1223 TCATGGCACTTGGCTTGTAGATGATTTTGTGGACATGCAACATCTTAGATTTCTTTT 1164
Db 956 euLeuThrValGlyLeuLeuValLeuLeuGluTyrGlnLeuAsnThrAlaGluAlaLeuP 976

QY 1163 TTCAAACTTTTAGTTTA-----TTAGATGTTACTCTTGACCATCCATCTTTACA 1113
Db 976 heLeuSerAlaSerValGlyLeuSerValAspLeuThrIleAsnTyrCysIleSerTyrH 996
QY 1112 GGGTTTCTTTTCTCCCATGGCTTCTAGGTCTCTCTTTCTGCCCCAGTAATTTCTC 1053
Db 996 isLeuCysProHisProAspArgLeuSerArgValAlaPheSerLeuArgGlnIleSerA 1016
QY 1052 AGTGT-----TTCTTTTCGGACTCATATTTCCCTAGGTGT 1017
Db 1016 rgAlaThrAlaMetThrThrGlyValLeuPheAlaSerGlyValIle----- 1031
QY 1016 GTCTTCTTTTCTCCCTGCGATCTT-----CTAGGTCTTCTTTGCTCG 973
Db 1032 -----MetLeuProSerThrIleLeuLeuTyrArgLysLeuGlyIleIleValMetM 1049
QY 972 CCCGTAATTTCTCAGATGTTTTTTCGAGGATCGTAATTTCTCAGGTGTTTTCTGCTG 913
Db 1049 etValLysPheLeuGlyCysGlyPheAla----- 1058
QY 912 ACGTTCGAACCTTCCAGATGCTCTTCTCCCAACTTGAATTTGTACAGATGTTT--- 860
Db 1059 -----SerPhePhePheGlnSerLeuCysCysPhePheGlyProG 1072
QY 859 -----CTCCACCAAGATGAGCCTCAT 838
Db 1072 lulyAsnCysGlyGlnIleLeuTyrProCysAlaHisLeuProTyrAspAlaGlyThrG 1092
QY 837 CAGGTGTTTTTCCACACAGCTTCCAGCTGTCAGGTCTCTTCCACAGGCTCAGC----- 782
Db 1092 luasp-----ProAspGluLysGlyArgAlaGly-----ProProGlyPheSerGluH 1108
QY 781 -----CTCATCAGGTGTTTTTCCACAGCTTTT----- 755
Db 1108 istyrGluLeuGlnProLeuAlaArgArgSerProSerPheAspThrSerThrAlaI 1128
QY 754 -----CAGCGTGTGAGGTCTTCTTCCCGAGGTGAGCTCATCAGTTT 709
Db 1128 hrSerLysLeuSerHisArgProSerIleLeuSerGluAspLeuGlnIleHisaspGlyS 1148
QY 708 CTGTCAGAGTT-----CCTTCTGATTTGGTATTTTGTGATTT 673
Db 1148 erCysCysLeuGlnHisAlaGlnAlaProValSerProArgAsp-----LeuLeuL 1165
QY 672 TTGATAATTTTCTGATATATTTTCAATTTGT----- 642
Db 1165 euAspHisGlnThrValPheSer--GlnCysProAlaLeuGlnThrSerSerProTyrLy 1184
QY 642 ----- 642
Db 1184 sGlnAlaGlyProThrProGlnThrThrIleArgGlnAspSerGlnGlyThrGl 1204
QY 641 -----TATGAATGTGAGAAATCCACAGTAACAGCATA 610
Db 1204 uProLeuGlnAlaLeuProGluGlyProAlaHisCysProLysSerLysValGlu----- 1222
QY 609 TGTTCGAGTTACTCCATATATCTGCAAAAGAGCTCACATTTTGTGTAAGAGCATGC 550
Db 1223 -----GluLeuProAspGlyLeuCysSerSerAlaSer-----ThrLeuGluGlyLe 1238
QY 549 AACTATCTCTGATGATCCAGACATACAGCAGCATGAGGTGTGTCATTTAACTTATTACG 490
Db 1238 uSerValSerAspAspThrCysAlaSerGluHisSer----- 1250
QY 489 CATTCGCAATTTGCAATTTTATCAGCAAAATTCACAAATTTGCTCCTCTTTCTGTTATG 430
Db 1251 -ValArgValProAspSerValGlyThrSerProGluValMetAsnGly----- 1266
QY 429 AGTAAAGTGGTGTGAGGCGCTTCTGCGACTTCGATCTGACACTGACACAGTGGACAGTT 370
Db 1267 -----ThrGlyHisProIleLeuGluArgGlyGlnLeuAs 1278

```

QY 369 TGGCACCACTGACAAAATCCACTATATAACGCATAATGGAAGCC-----AT 325
Db 1278 nGlyArgAspThrLeu---TrpLeuAlaLeuLysGluThrIleTyrAspProAsnMe 1297
QY 324 GTTGATACATACATCAGAGATTTATATCGGCACGAATCTATCAAAATATTGCAAGCCCT 265
Db 1297 tProAsnSerHisHis-----SerSerLe 1305
QY 264 CCTGTGCGCATGTAAG-----CCTTCATCAGA 238
Db 1305 uSerTrpLysGlyArgGlyGlyProGlyAspIleSerProValMetLeuProAsnSerGI 1325
QY 237 G-----GTGTCCTGTTCGCCAAGGACGTGAGCTGGCAGCTCTGCTAC 192
Db 1325 nProAspLeuProAspValTrpLeuArgArgProSerThrTyrThrSerGlyTyr 1343

RESULT 14
Q7YVS2
ID Q7YVS2 PRELIMINARY; PRT; 1679 AA.
AC Q7YVS2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypochemical protein.
GN TB927.2.1530.
OS Trypanosoma brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5691;
RN [1]
RF SEQUENCE FROM N.A.
RC STRAIN=GUT10.1;
RA El-Sayed N.M.A., Ghedin E., Song J., MacLeod A., Bringaud F.,
RA Larkin C., Wanless D., Peterson J., Hou L., Taylor S., Tweedie A.,
RA Biteau N., Khalak H.G., Lin X., Mason T., Hannick L., Caler E.,
RA Blandin G., Bartholomeu D., Simpson A.J., Kaul S., Zhao H., Bai G.,
RA Van Aken S., Uterback T., Haas B., Koo H.L., Unayam L., Suh B.,
RA Gerrard C., Leech V., Qi R., Zhou S., Schwartz D., Feldblyum T.,
RA Salzberg S., Tait A., Turner M.R., Ullu E., White O., Melville S.,
RA Adams M.D., Fraser C.M., Donelson J.E.;
RT "The sequence and analysis of Trypanosoma brucei chromosome II.";
RL Nucleic Acids Res. 0:0-0(2003).
DR EMBL; AB017167; AAQ15611.1; -.
KW Hypochemical protein.
SQ SEQUENCE 1679 AA; 197152 MW; 3A9D4F648B99F09A CRC64;

```

Alignment Scores:

```

Pred. No.:      0.000179      Length:      1679
Score:          140.00      Matches:      216
Percent Similarity: 31.11%      Conservative: 167
Best Local Similarity: 17.55%      Mismatches: 409
Query Match:      1.95%      Indels:      440
DB:              5          Gaps:      55

```

US-09-602-362E-22 (1-4115) x Q7YVS2 (1-1679)

```

QY 3849 CTGGCTGTGTTC-----TCCAAAGAGTTTCTACTGCTGTGTTCTCATGAGTTT 3799
Db 777 ValIleThrCysPheMetCysAsnSerMetSerLeuCysValLeuCysAlaLysAlaCys 796
QY 3798 CTGTTTCTCTCTCTCTTTTCATATGAATATACCGTTTAAATGGTATTGTAATTAAT 3739
Db 797 HisTyrValPheMetTyrLysSerMetSerLeuArgVal-----809
QY 3738 ATCTCCTCTTTTCTCTTTTAGAGATGTGTTCATTCCTCTCAAGAAATGAATATCAT 3679
Db 810 -----LeuCysAlaLysAlaCysHisTyrValPheTyrValGlnGlnHis 824
QY 3678 TGTATATC---TGCTTTGTGTGTCAGCTTTCTATGATGATGACTAATTCCTGTGAAGCCAC 3622
Db 825 -ValIleMetCysPheMetCysLys---SerMetSerLeuArgValTyrValGlnLysHi 843
QY 3621 AATTTTGTGTAGTGAATAATTTCTGACTAGAGACTCTGCTGTTTCAGGTGTTTG-- 3564
Db 3564

```

```

Db 843 sValIleThrCys-----PheMetTyrLysSerMetSerLeuArgValLeuCy 859
QY 3563 -----TTCCATTATCTT---GTGCTTTTGATACATGTGTTTCGCTTCTCTCATTTGA 3517
Db 859 sAlaLysAlaCysHisTyrValPheTyrValGlnGlnHisValIleThrCysPheMetCy 879
QY 3516 CACTCTTTTTCAGTTGTCTCTTTTGATGTT---CTGAAACCAATGTATTCTCTTAGAC 3460
Db 879 sAsn-----SerMetSerLeuCysAlaLysAlaCysHisTyrVal-----895
QY 3459 ATCTCCTCTAATTGAGATAATTTTAGTTTTCGATTTCCTTTGAGCTCAGAAAGTGT 3400
Db 896 -----PheTyrValGlnGlnHisValI 903
QY 3399 GATGGAGCACTCATTGTTAATATCGTACTCTCATCAACATTCATTCTTTTGCAAACA 3340
Db 903 eMet-----CysPheMetTyrLysSerMetSerLeuArgValLeuCysGlu-- 918
QY 3339 GCATCTCCTCAATGTGGAGCAGGTTCTTACTTTTCTTGATGTGCACAAATGATCATGGC 3280
Db 919 -----ThrAlaCysHisTyrValPheTyrValGlnLysHisValIleTh 933
QY 3279 TTGT-----ACAGCGAAGCCAGTCAGGATGTT 3253
Db 933 rCysPheMetTyrLysSerMetSerLeuArgValLeuCysAlaLysAlaCysHisTyrVa 953
QY 3252 GTATTCAATTTCTGCCTCTAGTTTCTTGTCTGTTTTCCTCAATTAGAAATGAGCAT 3193
Db 953 lPheTyrValGlnGlnHisValIleMetCysPheMetCysLysSerMetSerLeuArgVa 973
QY 3192 GTTTCTCAGCTATCAGAACTTTTAGCTGC-----3165
Db 973 lTyrValGlnLysHisValIleThrCysPheMetCysLysSerMetSerLeuArgValTy 993
QY 3164 -----CCACATATTGAGATCCCTTTTAGTAATGATTCCCTTTTCA-----GTTTT 3121
Db 993 rValGlnLysHisValIleThrCysPheMetCysAsnSerMetSerLeuCysValLeuCy 1013
QY 3120 AGGCTCATCTAAGTTCAGCATCTTTCTTTAAATCTTAAGTCTCTCAAGATTTATTTTC 3061
Db 1013 sAlaLysAlaCysHisTyrValPheMetTyrLysSerMetSerLeuArgValLeu---- 1031
QY 3060 CTTTTCCTGATGTTGTTTAGTGTGCTATTTCAGTCTTTTGCATGGCAATTTCTTTTCA 3001
Db 1032 -----CysAlaLysAlaCysHisTyrValPheTyrValGlnGl 1044
QY 3000 ACATGCAATTTTCATGAAGAATAATTTTCATTTTCATGAGTTGAGAA-----2955
Db 1044 nHisValIleMetCys-----PheMetCysAsnSerMetSerLeuCysVa 1059
QY 2954 -----ACCTATTCAAATTTCTTCTACACCTTCAATTCTTATCTTGTATCTGAGA 2905
Db 1059 lLeuCysAlaThrAlaCysHisTyrValPheTyrValGln-----Gl 1073
QY 2904 GCCTTTCAAGTTGTTTCTACTTCTTCTATGCTGCTCTCGATTCCTTCTAATTC 2845
Db 1073 nHisValIleMetCysPheMetCysLysSerMetSerLeuArgValSerCysThrLysAl 1093
QY 2844 TTCTTAATTTTCTTTTATATATATCGCATTTCTCTTCTCTCTTTCTTCTGTTTAAGTCA 2785
Db 1093 aCys-HisTyrValPheTyrValGlnLysHisValIleThrCysPheMetCysAsnSerM 1113
QY 2784 ATCTACACTCCAGACTCTTG-----TTCCCATTTTAACTTTTGTGTTCTCTA 2740
Db 1113 etSerLeuCysValLeuCysAlaLysAlaCysHisTyrValPheTyrValGlnGlnHisV 1133
QY 2739 CTGTGATTATTTCTTTTGTCTGAGAGTTCTTTTTCAGTA-----C 2698
Db 1133 alIleThrCysPheMetCysAsnSerMetSerLeuCysValLeuCysAlaLysAlaCysH 1153
QY 2697 AAAAACTCTTTTTCATTGTTCCATTTTCTGTACGTTTTCAGGTGTTTCAGGTGTTTGAAGTT 2638
Db 1153 is-----TyrValPheHisValGln-----1159

```


Db 1635 allileThrCysPheMetCysLysSerMetSerLeuArgValLeuCysThryLeuAlaCysH 1655
Qy 568 AT-----TTTGCTAAGAGCATGCACTATCTGTGATGATCCAGACATCAGCAGC 518
Db 1655 isTyValPheTyValGlnLysHisVal-IleThrCysPheMet-----TyrLysSer 1672
Qy 517 ATGAGGCTGTGCATT 503
Db 1673 MetSerLeuCysVal 1677
RESULT 15
Q8IIR3
ID Q8IIR3 PRELIMINARY; PRT; 1830 AA.
AC Q8IIR3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN PF11_0525.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chant M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Perten M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings J.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum";
RL Nature 419:498-511(2002).
DR EMBL; AS014837; AAN35690.1; -.
DR InterPro; IPR000048; IQ_region.
DR Pfam; PF00612; IQ; 1.
DR PROSITE; PS00096; IQ; 1.
KW Hypothetical protein.
SQ SEQUENCE 1830 AA; 218783 MW; 1C5BB15A4F423F44 CRC64;
Alignment Scores:
Pred. No.: Length: 1830
Score: 138.00 Matches: 179
Percent Similarity: 36.52% Conservative: 153
Best Local Similarity: 19.69% Mismatches: 332
Query Match: 1.95% Indels: 246
DB: Gaps: 40
US-09-602-362E-22 (1-4115) x Q8IIR3 (1-1830)
Qy 1376 ATCTGCTTCAAGCCGCACTTGAAGCAAACTTTTCCAAATAGCCTTTGAATGAAGA 1435
Db 797 lleCyslleLys-----LeuAsnSerLysAspIleCysAsnAsnValHisGluLeuMet 814
Qy 1436 ATG-----AAACAACATTGAAGCAGATCCGCTG 1462
Db 815 MetCysArgIleAsnLysGluLysLeuSerValAspGluLysHisLysLysGluLeu 834
Qy 1463 TTCCACCAAGATCCAAACAAGGACTATAAGAAATCTTGGGATCTTAGAGTCTCTGGA 1522
Db 835 TyrTyrIleAspGluGluHisAsnLeuProGluAspThrPheAsnAsnThrMetGluAsn 854
Qy 1523 ---GACTGTTCACAGAGGATGTGTTTACCAAGGCTACACATCAAAAAGAAATAGATAAA 1579
Db 855 ValAspAsnMetAspAsnAspGlnIleCysGluAsnAspValLysAspAsnMetGluGlu 874
Qy 1580 AAAATGGAAATAGA-----AGACTCCCTAATAAAGTGGTCTTCTGAAGCTACCTCGG 1633

Db 875 IlePheGlnAsnLysGlnAsnLysGlnAsnHisIleTyrPheAsnAsnLeuGlyArg 894
Qy 1634 AATGAAATTTCTATTCCACTAAAGCTAGAAATTTGAAGACA---TGCAAACTTTCAAGCG 1690
Db 895 AsnHisIleSerLeuIleAsn-ThrAsnAsnLeuLysAsnGluGluLysLeuGluArg 914
Qy 1691 AGCTCCGGGAAAGCCATCTGCTTCAGGCTCCACTGAAATCAAAAGTCTGTCCCAATAA 1750
Db 914 g-----LysValSerPheAlaSer----- 920
Qy 1751 GCCTTGGAAATGAAAAATGAAAAACATGGAGGAGATGAGAACTCCCATCAGNATCAAAAC 1810
Db 921 -----Th 921
Qy 1811 AAAGCACTATAAGAAATTTCTGGGATACTGGAGTCTCTGTAGACTGTTCACAGAAGAT 1870
Db 921 rlysglu-----LeuCysLeuLeuAsnAspAspIle 932
Qy 1871 TGTGTTTACCAGGCTGCGCTCAAAAAGATAGATAAAATATGGAATATAGAGGCTC 1930
Db 932 rValGluThrLysThrMetLeu-----LeuGluGlu 943
Qy 1931 CTTGTTAAAGTGTCTCTGAGGCTAACTGGGAATGAAAGTTCTATCCAACTAAAGCCT 1990
Db 943 rIleLeuLysHisValHisSerLys-----TyrAsnAsn----- 954
Qy 1991 AGAATTGATGATGACATGCAAACTTCAAGCAGCGCTCCGAGAGCCATTCGCTTCGAGCCT 2050
Db 955 -----LysCysMetAsnSerAsnIleIleTyrAspLysHisGlnSerIleHisTy 971
Qy 2051 CCATTGAAATCAAAAGTCTGCCAAATAAACTTGGGA---ATTGAAGATGACAAACATT 2107
Db 971 rAspAsnAsnLysLysSerLeuArgGluAsnSerGlySerMetArgAsnGluAsnLeu 991
Qy 2108 G-----AGGCAGATGAGAACTCCCATCAAAATCCAAACAAAGGAC 2146
Db 991 uPheIleHisAspAsnHisLysLysGluLysValGluProIleLysIleSerTyrAlaTh 1011
Qy 2147 TATGAGAAAGTCTCTGGGATCTCGAGTCTCTGTAGACTGTTCACAGAAGATGCTTT 2206
Db 1011 rThrArgAspLeuIleGly-----LysLysGluCysLe 1022
Qy 2207 ACCGAGCTACACATCAAAAGATAGATAAAATATGGAATAAGAGTCTCTCAT 2266
Db 1022 u-----LysGlyLysLeuGlu-ValGluAsnAsnAspValLysAsp 1036
Qy 2267 ATGAGTCTTCTGAGGCTCCCTGAGATGAAATGAAATGAAATGAAATGAAATGAAATG 2326
Db 1036 spGlu-----LysGlyAspAsnAsnValLysAspAspLysLysGlyAspAsn 1053
Qy 2327 ATGCAATCAAACTTTCAAAAGCAGGCTCCCGAGAGCCATCTGTCTCGAGCTGATTGA 2386
Db 1053 snAsnValLysAspAspLysLysGly----- 1061
Qy 2387 AATGAAAGTCTGTTCCAAATAAACCTTGGAAATTAAGAATGAACAACATTGACAGATC 2446
Db 1062 -----AspAsnAsnAsnValLysAspG 1070
Qy 2447 AGATTTCCCTTCAGAT-----CAAAACAAGAGCTTCAGAAATCTTGGGATTC 2497
Db 1070 lLysGlyAspAsnAsnAsnAspValLysAspGluLysArgAspAsnAsnAsnVal 1090
Qy 2498 TGAAGTCTCCGTGAGCTGTTTTCACAGAAGGATGCTGTGTACCCAGGCTACACACAAAAG 2557
Db 1090 allyAsp-----AspGluAsnAspAsnMet-AspAsnVal-----Lys 1102
Qy 2558 AAAGGATAAATAGTGAAGATTTAGAGATCTACTAGCCTATCAAAATCTTGTACAGTT 2617
Db 1103 ArgValTyrAsnAsnAspLysGluThrThrLeuAsnTyrLysAsnLeuLeuValAsn 1122
Qy 2618 CATCTTGTG---AAAGGCAAGGGA----- 2638
|||||:|||||

Db 1123 HisValIleAlaLysGlyLysGluIleAsnAsnAspLysAsnArgAsnGluAspTyrTyr 1142
QY 2639 ACTTCAAAAGACACTGTGAACACGTACAGGAAATGGAAACATGAAAGAAAGTTTGTG 2698
Db 1143 ThrTyrLysAspAsnThrGluLysLysAsnGluLysGluHisMetLeuAsnAsnTyrIle 1162
QY 2699 TACTCAAAAGAACTGTGACAGCAAAAGAAATAAATCACAGTAGAGAACCAAAAGTTAAA 2758
Db 1163 HisAspLysGluLysMetLysIleLeasp-----AsnGluAspAsnThrAsn 1178
QY 2759 TGGGAACAAGACTGTGACGTAGATTGACTTTAAACCAAGAAAGAGAGAGAGAAATGC 2818
Db 1179 AsnLysGluAspThrProIleAsnIle-----TyrLysLysLysAsnAspAsnIle 1195
QY 2819 CGATATATTAAAGAAATAATTAGGAAGAAATAGGAAGAAATCGAGAGCAGCATAGAAAGAGT 2878
Db 1196 TyrIleSerLysGluCysIleThrAsnAsnAsnAsnAsnAsnAsnAsnAsn--- 1214
QY 2879 TAGAAGTGAACACAACTTGAAGGCTCTCAGATACAAAGATAAGAAATTGAAGGTGAGAA 2938
Db 1215 -----AspAsnAsnGlu 1218
QY 2939 AGTAATTTGAATAGTTTCTCAACTCATGAATGAAATATTATTCTTACATGAAATTCAT 2998
Db 1219 AsnAsnGluAsnAsnGluAsnAsnGluAsnAsnSerTyrTyrAsnAsnGlu 1238
QY 2999 GTTGAAGAAAGATTGCCATCAAACTGGAATGACACACTAAACACCAATCCAGGAAA 3058
Db 1239 GlyGluLysAsnThrTyrAsnGlnAsnIleSerAsnHisLeuGlnAsnLysAspLysLys 1258
QY 3059 AG-----GAAATAAATCTTTGAGGACTTAAGATTTTAA-----GAA 3096
Db 1259 LysTyrThrAsnSerIleGluAsnGluAlaLeuIleLysAsnAsnIleLysAsnAspGlu 1278
QY 3097 AAGATGCTGAACCTAGATGACCCCTAAACT-----GAAAGGAATCATTA 3141
Db 1279 LysTyrIleAsnTyrGlnAsnTyrLysGluArgLysIleLeuHisLysAsnAsnMet 1298
QY 3142 CTA-----AAGGATCTCAATATGTGGCAGCTAAAGTTCTGATGATGAGAAACAAT 3195
Db 1299 IleAsnLysAsnAspIleGluSerThrProGlnLeuGlnMetSerTyrIleLeuAsnAsn 1318
QY 3196 GCTCATTTAAATGAGGAAAGACAGACAAAGAAACTAGAGCAGAAATTTGAATACACC 3255
Db 1319 SerGlnAsnAsnTyrSerIleAsnLysAspAsnIleIleLysSerProIleGlu----- 1336
QY 3256 ATCCTGACTGCTTCGGCTGTACAGCCATGATCAATTTGACATCAAGAAAGTAAGAA 3315
Db 1337 -----IleValArgLysLeuLeuAsnPheAspGluLysGlnLys----- 1349
QY 3316 CTGCTTCCACATGAGAGATGCTGTTTGCAGAAATGAATGTTGATGTGAGAGTAC 3375
Db 1350 -----AsnIleGluGluCysIle-----GluAsn 1357
QY 3376 GATATTAAACATGATGCTCCATCACCCTTCTGAGCTCAAAGGAAATCCAAACCTAA 3435
Db 1358 IleAsnAsnAsnAsnAsnAsnIleAsn-----AspLys-TyrProLysGluLys 1373
QY 3436 AAATTAATCTCAATTAGCAGGAGATGCTTAAGAGAAATACATTTGGTTTCAGAACATCAAA 3495
Db 1373 sAsnLysLeuLysGluAspGlnIle-----LysHis-LeuL 1385
QY 3496 AGAGACAACGTGAAAGAGTGTCAATGAAGGAAGCAACACATGATCAAAAGCAGCAAGA 3555
Db 1395 yTyrAspAspMetAsnSerMetProAsnAspAsnAspAsnValHisHisAsnLeuAsnA 1405
QY 3556 TAATGGAACAACACCTGAACAGCAGAGTCTTAGTCAGAAATTTATTCAACTACAGCA 3615
Db 1405 snLeuCysMetIleSerLysGluGluAsnIleIleThrLys---LysAsnThrThrGluL 1424
QY 3616 AAAATTTGGCTTCCACAGCAATTAGTCATCAGTAAAGAGCTGCAACAAAGACAGATA 3675
Db 1424 yTyrCysAsnIle-----PheLysAsnIleAsnAsnLysIle-----AsnCysAsnA 1440

QY 3676 ACAATGATATTTCATTTCTTGAGAGGAATGCAACACATCTCTCTAAAGAGAAA----- 3728
Db 1440 snAsnGluAspAsnTyrLeuIle---AsnArgAsnGluLysLysGluLysSerLysI 1459
QY 3729 -----AAGAGGAGATATAATTAATTAACATACCATTTAAACACCGTATATTCAATAT 3777
Db 1459 leAsnLysAsnLysAsnAsnTyrAsnLysGluTyrAsnAspGluGlu----- 1474
QY 3778 GAAAAAGAGAAAGAGAAACAGAAACTCATCAGAACAGCAGTAGAGAACTTCTTTGGAGAA 3837
Db 1475 --LysLysAspLysAspHisIleAsnHisAsnIleSerAspAsnPheIleMetLysA 1494
QY 3838 AACACAGACAGACTTTTACTCACACTCTAGAGCCAGCTCTAGCATCACCCTATGTTG 3897
Db 1494 snAsnAsnIleLysAsnValGluGluAsnGlnAsnArgGlnAspValValValAspValG 1514
QY 3898 AAAATCTTCAAGATC 3914
Db 1514 lyAsnIleIleSerIle 1519

Search completed: July 15, 2004, 09:25:56
Job time : 308.218 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 08:45:37 ; Search time 34.7876 Seconds
(without alignments)
12318.669 Million cell updates/sec

Title: US-09-602-362E-22
Perfect score: 7086
Sequence: 1 ctatgctatacagcaacac.....gagtggaactccactggaaa 4115

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+42p.model -DEV=xlp
-Q=/Cgn2_1/USPTO_spool_p/US09602362/runat_15072004_093625_21988/app_query.fasta_1.10325
-DB=SwissProt_42 -QFMT=fastan -SUFFIX=resp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptlo -NORM=ext -HRAPSZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09602362 @CGN_1_1_134 @runat_15072004_093625_21988 -NCPU=6 -ICPU=3
-NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------------|--------------------|
| 1 | 144 | 2.0 | 3911 | 1 AKA9 HUMAN | Q99996 h a-kinase |
| 2 | 139 | 2.0 | 1818 | 1 HMW2 MYCPN | P75471 mycoplasma |
| 3 | 138 | 1.9 | 590 | 1 NU5M TRYBB | P04540 trypanosoma |
| 4 | 136 | 1.9 | 976 | 1 SCPI HUMAN | Q15431 homo sapien |
| 5 | 133.5 | 1.9 | 527 | 1 NU5M CAEEL | P24896 caenorhabdi |
| 6 | 130.5 | 1.8 | 1433 | 1 REST CHICK | Q42184 gallus gall |
| 7 | 130 | 1.8 | 547 | 1 NU5M ASCSU | P24884 ascaris suu |
| 8 | 130 | 1.8 | 1163 | 1 SECC CLOAB | Q97fk1 clostridium |
| 9 | 130 | 1.8 | 2230 | 1 GOA4 HUMAN | Q13439 homo sapien |
| 10 | 130 | 1.8 | 2663 | 1 CENE HUMAN | Q02224 homo sapien |
| 11 | 127 | 1.8 | 1005 | 1 RA50 METJA | Q58718 methanococc |
| 12 | 125.5 | 1.8 | 2245 | 1 MYSJ DICTDI | P54697 dictyosteli |
| 13 | 125 | 1.8 | 1928 | 1 MYS1 YEAST | P08964 saccharomyc |
| 14 | 123.5 | 1.7 | 3259 | 1 GOB1 HUMAN | Q14789 homo sapien |
| 15 | 123 | 1.7 | 313 | 1 NU2M RHISA | Q99817 rhiocephal |
| 16 | 121.5 | 1.7 | 1411 | 1 FEAL HUMAN | Q15075 homo sapien |
| 17 | 120.5 | 1.7 | 409 | 1 NU4M CAEEL | P24892 caenorhabdi |
| 18 | 120.5 | 1.7 | 1972 | 1 MYHB_RABIT | P35748 oryctolagus |

ALIGNMENTS

RESULT 1

AKA9 HUMAN
ID AKA9 HUMAN STANDARD; PRT; 3911 AA.
AC Q99996; O14869; O43355; O94895; Q9UQH3; Q9UQ04; Q9Y6B8; Q9Y6Y2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE A-kinase anchor protein 9 (Protein kinase A anchoring protein 9)
DE (PRK9) (A-kinase anchor protein 450 kDa) (AKAP 450) (A-kinase anchor
DE protein 350 kDa) (AKAP 350) (hgAKAP 350) (AKAP 120 like protein)
DE (Hyperion protein) (Yotiao protein) (Centrosome- and Golgi-localized
DE PKN-associated protein) (CG-NAP).
GN AKAP9 OR AKAP450 OR AKAP350 OR KIAA0803.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;
RX MEDLINE=98151389; PubMed=9482789;
RA Lin J.W., Myszyński M., Madhavan R., Sealock R., Kim J.U., Sheng M.;
RA "Yotiao, a novel protein of neuromuscular junction and brain that
RA interacts with specific splice variants of NMDA receptor subunit
RA NR1.";
RL J. Neurosci. 18:2017-2027(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.
RX MEDLINE=99219864; PubMed=10202149;
RA Witczak O., Skallhegg B.S., Keryer G., Bornens M., Tasken K.,
RA Jahnsen T., Oerstavik S.;
RA "Cloning and characterization of a cDNA encoding an A-kinase anchoring
RA protein located in the centrosome, AKAP450.";
RL EMBO J. 18:1858-1868(1999).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Brain;
RX MEDLINE=99287934; PubMed=10358086;
RA Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.;
RA "Characterization of a novel giant scaffolding protein, CG-NAP, that
RA anchors multiple signaling enzymes to centrosome and the Golgi
RA apparatus.";

C 19 118.5 1.6 615 1 NUOL_BUCAP
C 20 118.5 1.6 634 1 YCX3_EUGR
C 21 118 1.6 486 1 NUON_BUCAP
C 22 118 1.7 978 1 RA50_AQUAE
C 23 118 1.7 1727 1 ALM1_SCHPO
C 24 117.5 1.7 1290 1 RA50_SCHPO
C 25 116 1.6 1938 1 MYHD_HUMAN
C 26 115.5 1.6 469 1 NUON_BUCAI
C 27 115.5 1.6 3210 1 CENF_HUMAN
C 28 114 1.6 1957 1 SPOF_SCHPO
C 29 113 1.6 494 1 NUON_BUCBP
C 30 111.5 1.6 1790 1 USOI_YEAST
C 31 111 1.6 1961 1 MYH9_RAT
C 32 110.5 1.5 444 1 NU4M_LOCM1
C 33 110.5 1.6 845 1 SCPI_MESAU
C 34 110.5 1.6 1875 1 MLP1_YEAST
C 35 110 1.5 284 1 COX3_LEITA
C 36 110 1.6 1427 1 REST_HUMAN
C 37 110 1.6 1934 1 MYH7_MESAU
C 38 109.5 1.5 310 1 Y160_BUCAP
C 39 109 1.5 527 1 NU2M_ACACA
C 40 109 1.5 1505 1 SCP2_RAT
C 41 109 1.5 1679 1 Y105_YEAST
C 42 109 1.5 1935 1 MYH7_PIG
C 43 108.5 1.5 229 1 ATP6_TRYBB
C 44 108.5 1.5 736 1 NU5C_OENHO
C 45 108.5 1.5 1530 1 SCP2_HUMAN

Q8K9X7 buchnera ap
P31916 euglena gra
Q8K9X5 buchnera ap
O67124 aquifex aeo
Q9utk5 schizosacch
Q9utj8 schizosacch
Q9ukx3 homo sapien
P57264 buchnera ap
P49454 homo sapien
Q10411 schizosacch
Q89at4 buchnera ap
P25386 saccharomyc
Q62812 rattus norv
Q62812 lucusta mig
Q65424 mesocricetu
Q02455 saccharomyc
P14546 leishmania
P30622 homo sapien
P13540 mesocricetu
Q8K9X4 buchnera ap
Q37376 acanthamoeb
Q70608 rattus norv
P40457 saccharomyc
P79293 sus scrofa
P24499 trypanosoma
Q9mt14 cenothera h
Q9bx26 homo sapien

RL J. Biol. Chem. 274:17267-17274 (1999).
 RN [4]
 RA SEQUENCE FROM N.A. (ISOFORM 1).
 RX Kemmer W.A., Deiss S., Schwarz U.;
 RT "Cloning of Hyperion";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RA SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).
 RX TISSUE=Gaetac parietal cell;
 RX MEDLINE=99115654; PubMed=9915945;
 RA Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,
 RA Trotter K.W., Milgram S.L., Goldenring J.R.;
 RT "AKAP350, a multiply spliced protein kinase A-anchoring protein
 associated with centrosomes.";
 RL J. Biol. Chem. 274:3055-3066 (1999).
 RN [6]
 RA SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).
 RX TISSUE=Lymphoblast;
 RA Hinds K., Sutter C., Becker M., Hawkins M.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RA SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).
 RX TISSUE=Lung;
 RA Milgram S.L., Goldenring J.R., Schmidt P.H.;
 RT "AKAP350: A multiply spliced family of proteins with centrosomal
 association.";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RA SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).
 RX TISSUE=Brain;
 RX MEDLINE=99087487; PubMed=9872452;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XI.
 The complete sequences of 100 new cDNA clones from brain which code
 for large proteins in vitro.";
 RL DNA Res. 5:277-286 (1998).
 RN [9]
 RA SEQUENCE OF 17-1800 FROM N.A.
 RA Wu X., Graves T., Bradshaw H.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Binds to type II regulatory subunits of protein kinase
 A. Scaffolding protein that assembles several protein kinases and
 phosphatases on centrosome and Golgi apparatus where physiological
 events can be regulated by phosphorylation state of protein.
 CC substrates. Isoform 4/Yotiao is associated with the N-methyl-D-
 aspartate receptor and is specifically found in the neuromuscular
 junction (NMJ) as well as in neuronal synapses explaining that its
 role may be to organize postsynaptic specializations.
 CC -!- SUBUNIT: Interacts with the regulatory region of protein kinase N
 (PKN), protein phosphatase 2A (PP2A), protein phosphatase 1 (PP1)
 and the immature non-phosphorylated form of PKC epsilon.
 CC -!- SUBCELLULAR LOCATION: Centrosomal in many cell types and
 cytoplasmic in parietal cells.
 CC -!- Event=Alternative splicing; Named isoforms=6;
 CC Name=1;
 CC IsoId=Q99996-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q99996-2; Sequence=VSP_004102, VSP_004107;
 CC Name=3; Synonyms=CG-NAP;
 CC IsoId=Q99996-3; Sequence=VSP_004102, VSP_004105, VSP_004107;
 CC Name=4; Synonyms=Yotiao;
 CC IsoId=Q99996-4; Sequence=VSP_004103, VSP_004104;
 CC Name=5;
 CC IsoId=Q99996-5; Sequence=VSP_004108;
 CC Name=6; Synonyms=AKAP350;
 CC IsoId=Q99996-6; Sequence=VSP_004106, VSP_004107, VSP_004109;
 CC -!- TISSUE SPECIFICITY: Widely expressed, Isoform 4/Yotiao is highly
 expressed in skeletal muscle and in pancreas.
 CC -!- DOMAIN: RII-binding site, predicted to form an amphipathic helix,
 could participate in protein-protein interactions with a
 complementary surface on the R-subunit dimer.

CC -!- CAUTION: Ref.6 sequence differs from that shown due to two
 frameshifts in positions 3782 and 3811.
 CC -!- CAUTION: Ref.9 sequence differs from that shown due to four
 frameshifts in positions 29, 1653, 1699 and 1735.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ131693; CAB40713.1; -;
 DR EMBL; AB019691; BAA78718.1; -;
 DR EMBL; AJ010770; CAA09361.1; -;
 DR EMBL; AF026245; AAB86384.1; -;
 DR EMBL; AF083037; AAD22767.1; -;
 DR EMBL; AC004013; AAB96867.1; ALT_FRAME.
 DR EMBL; AF091711; AAD39719.1; -;
 DR EMBL; AF018346; BAA34523.1; -;
 DR EMBL; AC000666; AAC60380.1; ALT_FRAME.
 DR Genew; HGNC:379; AKAP9.
 DR MIM; 604001; -;
 DR GO; GO:0005813; C:centrosome; TAS.
 DR GO; GO:0005856; C:cytoskeleton; TAS.
 DR GO; GO:0004973; F:N-methyl-D-aspartate receptor-associated pr...; TAS.
 DR GO; GO:0005515; F:protein binding; TAS.
 DR GO; GO:0007165; P:signal transduction; TAS.
 DR GO; GO:0007268; P:synaptic transmission; TAS.
 DR GO; GO:0006810; P:transport; TAS.
 KW Coiled coil; Alternative splicing; Polymorphism.
 FT DOMAIN 2554 2567 PKA-RII SUBUNIT BINDING DOMAIN.
 FT DOMAIN 164 914 COILED COIL (POTENTIAL).
 FT DOMAIN 944 1022 COILED COIL (POTENTIAL).
 FT DOMAIN 1100 1185 COILED COIL (POTENTIAL).
 FT DOMAIN 1253 1280 COILED COIL (POTENTIAL).
 FT DOMAIN 1336 1392 COILED COIL (POTENTIAL).
 FT DOMAIN 1434 1459 COILED COIL (POTENTIAL).
 FT DOMAIN 1585 1659 COILED COIL (POTENTIAL).
 FT DOMAIN 1857 2455 COILED COIL (POTENTIAL).
 FT DOMAIN 2544 2561 COILED COIL (POTENTIAL).
 FT DOMAIN 2603 2776 COILED COIL (POTENTIAL).
 FT DOMAIN 3065 3092 COILED COIL (POTENTIAL).
 FT DOMAIN 3124 3470 COILED COIL (POTENTIAL).
 FT DOMAIN 3587 3699 COILED COIL (POTENTIAL).
 FT DOMAIN 3726 3730 POLY-LEU.
 FT DOMAIN 203 292 GLN-RICH.
 FT DOMAIN 321 1010 GLU-RICH.
 FT DOMAIN 1846 2772 Missing (in isoform 2 and isoform 3).
 FT VARSPLIC 17 28 /FTID=VSP_004102.
 FT VARSPLIC 1637 1642 CLOEEL -> LATRED (in isoform 4).
 FT VARSPLIC 1643 3911 Missing (in isoform 4).
 FT VARSPLIC 2175 2182 Missing (in isoform 3).
 FT VARSPLIC 2175 2183 SADTFQKVE -> Q (in isoform 6).
 FT VARSPLIC 2175 2183 /FTID=VSP_004106.
 FT VARSPLIC 2895 2907 VTFYFNMCFSTLC -> GSSPELAHSDAVOTREICSS
 (in isoform 2, isoform 3 and isoform 6).
 FT VARSPLIC 2895 2948 /FTID=VSP_004107.
 FT VARSPLIC 3901 3911 Missing (in isoform 5).
 FT VARSPLIC 3901 3911 Missing (in isoform 6).
 FT VARSPLIC 3901 3911 STTQFHAGMRR -> ALSLTTSWQHHSARPTAPLFFELSH
 SLG (in isoform 6).
 FT VARIANT 1347 1347 /FTID=VSP_004109.
 FT VARIANT 1347 1347 K -> Q.
 FT CONFLICT 76 76 E -> Q (IN REF. 3).
 FT CONFLICT 475 475 M -> I (IN REF. 3).
 FT CONFLICT 554 554 E -> G (IN REF. 3).

| | | | | | |
|----|----------|------|------|-------------------------------|--|
| FT | CONFLICT | 638 | 538 | R -> S (IN REF. 3). | |
| FT | CONFLICT | 663 | 663 | N -> S (IN REF. 3). | |
| FT | CONFLICT | 913 | 913 | H -> N (IN REF. 3). | |
| FT | CONFLICT | 956 | 956 | K -> N (IN REF. 3). | |
| FT | CONFLICT | 980 | 982 | OKH -> PKP (IN REF. 1 AND 2). | |
| FT | CONFLICT | 997 | 997 | Q -> P (IN REF. 1 AND 2). | |
| FT | CONFLICT | 1001 | 1001 | Q -> P (IN REF. 1 AND 2). | |
| FT | CONFLICT | 1020 | 1020 | N -> D (IN REF. 3). | |
| FT | CONFLICT | 1028 | 1028 | V -> E (IN REF. 3). | |
| FT | CONFLICT | 1626 | 1626 | R -> P (IN REF. 1 AND 2). | |
| FT | CONFLICT | 1703 | 1703 | N -> T (IN REF. 3). | |
| FT | CONFLICT | 1707 | 1707 | V -> G (IN REF. 3). | |
| FT | CONFLICT | 1802 | 1803 | MISSING (IN REF. 5). | |
| FT | CONFLICT | 1843 | 1843 | A -> P (IN REF. 3). | |

| | | | | |
|------------------------|----------|---------------|------|--|
| Alignment Scores: | | | | |
| Pred. No.: | 0.000178 | Length: | 3911 | |
| Score: | 144.00 | Matches: | 175 | |
| Percent Similarity: | 39.36% | Conservative: | 143 | |
| Best Local Similarity: | 21.66% | Mismatches: | 252 | |
| Query Match: | 2.03% | Indels: | 240 | |
| DB: | 1 | Gaps: | 39 | |

US-09-602-362E-22 (1-4115) x AK9_HUMAN (1-3911)

| | | | |
|----|------|--|------|
| Qy | 1989 | CTAGAATTGATGACATGCACAACTCTCAAGCAGGCTCCCGAGAGCCATCTGCCTTCGAGC | 2048 |
| Db | 169 | LeuGluMetMet----- | 179 |
| Qy | 2049 | CTCATTGAATCAAAAGTCGTGTCCAATAAACCTTGGAAATTAGAATGAACAAACATTC | 2108 |
| Db | 180 | GlnHis-GluIleGluLeuAenA-gGluLeuGluGluMetArgVal----- | 195 |
| Qy | 2109 | AGGCAGATGAGAATCCCATCAATCAAAACAAAGGACTATGAA--AGTCTTGGGA | 2165 |
| Db | 196 | -----ThrTyrGlyThrGluGlyLeuGlnGluGluGluPheGln | 209 |
| Qy | 2166 | TTCTGGAGTCCTCTAGACTGTTTCCAGAAGGATGGTGTATTACCGAGCTACACATCAA | 2225 |
| Db | 209 | uAlaAlaIleLys-----GlnA-gAspGlyIleIleThrGlnLeuThrAla-- | 224 |
| Qy | 2226 | AAGATAGATAAATAATGGAATAAGAGAGTCTCTGATAAT----- | 2268 |
| Db | 225 | -----AsnLeuGlnGlnAlaA-gA-gGluLysAspGluThrMetArgGluPheLe | 241 |
| Qy | 2269 | -GAGGTTTCTGAAGGCTCCCTGAGATGAAGATTCATTATCCACTAAAGCCTTGAATTGA | 2327 |
| Db | 241 | uGluLeu-ThrGluGlnSerGlnLysLeuGlnIleGlnPheGln-----GlnLeuG | 258 |
| Qy | 2328 | TGGCATGCAACATTTCAAAGCAGGCTCCCGAGAGCCATCTGCTTCGAGCCTCATGAA | 2387 |
| Db | 258 | lnAlaSerGluThrLeuArgAsnSerThrHisSerSerThrAlaAlaAspLeu-LeuGln | 277 |
| Qy | 2388 | ATGAAAAGTCTGTTCCTCAATAAACCCTTGGAAATTAAGAATCAACAAACATTGAGACATCA | 2447 |
| Db | 278 | AlaLysGln-----GlnIle-----LeuThrHisGlnGlnGlnLeuGluGln | 292 |
| Qy | 2448 | GAITTCCTTCAGATCAAAACAAAGAGGTTTGAGAAATCTGGGATTCGAAGTCCTC | 2507 |
| Db | 293 | AspHisLeuLeuGluAspTyrGlnLysLysGluAspPhe-----ThrMetGln | 309 |
| Qy | 2508 | GTGGACTGTTTCACAGAGGATGGTGTTGACCAGGCTACACACAAAGAGGATAAA | 2567 |
| Db | 310 | IleSerPheLeuGlnGlu-----LysIleLysValTyrGluMetGluGlnAspLys | 326 |
| Qy | 2568 | ATAGTGAATAATTAGAGATTCACCTACGCTATCAAAATCTGGTACAGTTCATCTGTGA | 2627 |
| Db | 327 | LysValGluAsnSerAsn----- | 332 |
| Qy | 2628 | AAGCAAGGGAACCTTCAAAAGACACTGTG-----NACAGTACAGGAAAATG | 2675 |
| Db | 333 | LysGluGluLeuGlnLysGluThrIleLeuGluLeuAsnThr---LysIleIle | 351 |

Db 686 eSerGlnLysIleGluThrMetGlnPheGluLysAspAsnLeuIle-ThrLysGlnAsn 705
 QY 3574 AACACGAGTCTCTAGTCAGAAATTTTCAACTACAGCAAAAATTTGTG-----GCT 3627
 Db 706 GlnLeuIleLeuGluIleSerLysLeuLysAspLeuGlnSerLeuValAsnSerLys 725
 QY 3628 TCACAGCAATAGTCATCAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 3687
 Db 726 SerGluGluMetThrLeuGlnIleAsnGlu---LeuGlnLysGluIleGluIleLeu--- 743
 QY 3688 ATTCTTTCAGAGGAAATGCAACATCTCTCTAAAGAGAGAGAGAGAGATATTATTAC 3747
 Db 743 ----- 743
 QY 3748 AATACCATTTAAACCGTATATTCAATATGAAAGAGAGAGAGAGAGAGAGAGAGAG 3807
 Db 744 -----ArgGlnGluGluLysGluLys----- 750
 QY 3808 AGAACACGAGTACAGAACTCTTTGGAGAAACACAGACAGACTTTTACTCAGACTCATG 3867
 Db 751 -----GlyThrLeuGluGlnGluValGlnGluLeuGln 761
 QY 3868 CTAGAGCCAGCTAGCATCATCTATCTGAAATCTTACCAAGTCTGTGTACAGAT 3927
 Db 762 LeuLysThrGluLeuLysGlnMetLys-----GluLysGluAsn 776
 QY 3928 ACTATTTTGAAGAAATTCATGTTCTCTGAGCCTACAGCTACAGTAAATAAAGTGCA 3987
 Db 777 AspLeuGlnGlu---LysPheAlaGlnLeuGluAlaGluAsnSerIleLeuLysAspGlu 796
 QY 3988 AGATTACTTGTTCGGAATTCGATAGCTGACAGGATTCCTATCCCTGCTGTCAGCA 4047
 Db 796 YLysThrLeuGluAspMetLeuLysIleHis---ThrProValSer---GlnGluGlu 814
 QY 4048 GACTCATTCATCAACAGAAATCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4107
 Db 814 rgLeuIlePheLeuAspSerIleLysSerLysSerLys-----AspSerV 829
 QY 4108 CCTGAAA 4115
 Db 829 alTrpGlu 831

RESULT 2
 ID_HMW2_MYCPN STANDARD; PRT; 1818 AA.
 AC P75471.
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cytadherence high molecular weight protein 2 (Cytadherence accessory
 protein 2).
 DE HMW2 OR MPN310 OR MP526.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_taxID=2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29342 / M129;
 RA MEDLINE=97105885; PubMed=8948633;
 RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
 Herimann R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 pneumoniae.";
 RL Nucleic Acids Res. 24:4420-4449 (1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29342 / M129;
 RA MEDLINE=97252497; PubMed=9098065;
 RA Krause D.C., Proft T., Hedreya C.T., Hilbert H., Plagens H.,
 Herimann R.;
 RT "Transposon mutagenesis reinforces the correlation between Mycoplasma
 pneumoniae cytoskeletal protein HMW2 and cytodherence.";

RL J. Bacteriol. 179:2668-2677 (1997).
 CC -FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
 CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
 CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
 CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHERESIN PROTEINS
 CC IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
 CC SIMILARITY).
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB000051; AAB96174.1; -;
 CC EMBL; US9896; AAB52527.1; -;
 CC PIR; S73852; S73852.
 CC KW Phosphite; P75471; -;
 CC FT DOMAIN 31 880 COILED COIL (POTENTIAL).
 CC FT CYTADHERENCE; Structural protein; Coiled coil; Complete proteome.
 CC FT DOMAIN 919 1607 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1644 1755 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1786 1817 COILED COIL (POTENTIAL).
 CC SQ SEQUENCE 1818 AA; 215622 MW; 66DF4B08F0FCBC0 CRC64;

 Alignment Scores:
 Pred. No.: 0.000442 Length: 1818
 Score: 139.00 Matches: 236
 Percent Similarity: 35.48% Conservative: 200
 Best Local Similarity: 19.20% Mismatches: 384
 Query Match: 1.96% Indels: 409
 DB: 56 Gaps:

 US-09-602-362E-22 (1-4115) x HMW2_MYCPN (1-1818)
 QY 901 AAGTTCGACGTCAGCAGAGAAACACCTAGGAAATACGATCTCTGCAAAAAACATCTGA 960
 Db 506 LysLysGlnLysGlnAspGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 525
 QY 961 GAAATTTACGGCCAGCAAG 1020
 Db 526 GlnTyr-----GlnAlaAspPheGluAsnGluIleGluGluLysGlnAsnGlu 541
 QY 1021 CTAGGGAATATGAGTCCGAGAAAGAAACACCTAGGAAATACGATCTCTGCAAAAAAGGAGA 1080
 Db 542 Leu-----PheAlaSerGlnLysSerLeuGlnLysSerPheThrGlnLysLysAsnLys 559
 QY 1081 CTAGGAG 1134
 Db 560 GluAlaGluLeuAsn---GlnLysAlaGlnLysIleAlaGluAspTrpAlaHisLeuLys 578
 QY 1135 AGTAAACATCTAATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1188
 Db 579 GlnAsn-----LysHisHisAlaAspLeuGluLysPheLeu 591
 QY 1189 ---CCACAAAGAGATCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1245
 Db 592 GluGluGluPheAsnHisLeuGlnGlnLysHis-----LysLeuLeuGluAlaArg 609
 QY 1246 ACAAG 1305
 Db 610 ThrGlnPheAspAsnArgValSerLeuLeu-----SerAlaArgPhe 623
 QY 1306 AAGTGTATACCTGAGTCTATATATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1365
 Db 623 ----- 623
 QY 1366 CCTAAAAGCCATCTGCTTCAAGCCGATTTGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1425
 Db 624 -----LysGlnLysGlnAlaGluLeu----- 630

| | | | |
|---------------------------|--|---|------|
| Qy | 2732 | TTATTCCTTTTGTTCTGACAGTGTCTTTTTTCAGTAGCACAAAACCTCTTTTTCATTGTT----- | 267 |
| Dd | 522 | hetYrPhePheAsnIleAspCysIleMetLeuPheTrpArgIlePhePheValIleIleI | 542 |
| Qy | 2676 | -----CCATTTCCTGGTAGCGTGTTCACAGTGCCTTTTTGAAGTTCCT | 2634 |
| Dd | 542 | leLeuValValPheMetIlePheCysCysTrpTyRPhEValCysMeIleIlePheMetL | 562 |
| Qy | 2633 | TGCCTTTCAAGAAGTAACCTGTACCAGATTTTGATAGGCTAGTGAATCTCTAAATTTTC | 2574 |
| Dd | 562 | eUleuPheValTrpAsnPheValIleTyRphe-----ArgTyra | 575 |
| Qy | 2573 | CACATATTTATCCTTCTCTTTTGTGTGCTAGCTGGGTA | 2536 |
| Dd | 575 | snleuLysTyRCysLeuPhePheCysIleLeutrpiile | 587 |
| RESULT 4 | | | |
| ID | SCPI HUMAN | STANDARD; PRT; 976 AA. | |
| AC | Q15431; O14963; | | |
| DT | 15-JUL-1998 (Rel. 36, Created) | | |
| DT | 15-JUN-1998 (Rel. 36, Last sequence update) | | |
| DT | 15-MAR-2004 (Rel. 43, Last annotation update) | | |
| DE | Synaptonemal complex protein 1 (SCP-1 protein). | | |
| DB | SYCP1 OR SCPI. | | |
| OS | Homo sapiens (Human). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. | | |
| OX | NCBI_Taxid=9606; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RF | TISSUE=Testis; | | |
| RC | MEDLINE=97224467; PubMed=9119375; | | |
| RX | Meuwissen R.L.J., Meerts I., Hoovers J.M.N., Leschot N.J., | | |
| RA | Heyting C.; | | |
| EA | "human synaptonemal complex protein 1 (SCPI): isolation and | | |
| RT | characterization of the cDNA and chromosomal localization of the | | |
| RT | gene."; | | |
| RL | Genomics 39:377-384(1997). | | |
| RN | [2] | | |
| RP | SEQUENCE FROM N.A. | | |
| RF | TISSUE=Testis; | | |
| RC | MEDLINE=96037449; PubMed=9371398; | | |
| RX | Kondoh N., Nishina Y., Teuchida J., Koga M., Tanaka H., Uchida K., | | |
| RA | Inazawa J., Taketo M., Nozaki M., Nojima H., Matsumiya K., Namiki M., | | |
| OKuyama A., Nishimune Y.; | | | |
| RT | "Assignment of synaptonemal complex protein 1 (SCPI) to human | | |
| RT | chromosome lp13 by fluorescence in situ hybridization and its | | |
| RT | expression in the testis."; | | |
| RL | Cytogenet. Cell Genet. 78:103-104(1997). | | |
| CC | -I- FUNCTION: Major component of the transverse filaments of | | |
| CC | synaptonemal complexes (SCS), formed between homologous | | |
| CC | chromosomes during meiotic prophase. | | |
| CC | -I- SUBCELLULAR LOCATION: Nuclear. In tripartite segments of | | |
| CC | synaptonemal complexes, between lateral elements in the nucleus. | | |
| CC | Found only where the chromosome cores are synapsed. Its N-terminus | | |
| CC | is found towards the centre of the synaptonemal complex while the | | |
| CC | C-terminus extends well into the lateral domain of the | | |
| CC | synaptonemal complex (By similarity). | | |
| CC | -I- TISSUE SPECIFICITY: Testis. | | |
| CC | -I- DOMAIN: Consists of an alpha-helical stretch of 700 AA residues, | | |
| CC | flanked by N- and C-terminal globular domains. The C-terminal | | |
| CC | domain has DNA-binding capacity (By similarity). | | |
| CC | ----- | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | |
| CC | use by non-profit institutions as long as its content is in no way | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | |
| CC | or send an email to licensesib-sib.ch). | | |
| CC | ----- | | |
| DR | EMBL; X95654; CAA64956.1; -- | | |

DR EMBL; D67035; BAA22586.1; --
 DR Genew; HGNC:11487; SYCP1.
 DR MIM; 602162; --
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0005715; C:synaptonemal complex; TAS.
 DR GO; GO:0003677; F:DNA binding; TAS.
 DR GO; GO:0007131; P:meiotic recombination; TAS.
 DR GO; GO:0007283; P:meiotic recombination; TAS.
 DR GO; GO:0007129; P:synapsis; TAS.
 DR InterPro; IPR008827; SCP-1.
 DR Pfam; PF05483; SCP-1; 1.
 DR Nuclear protein; Meiosis; Cell division; Phosphorylation;
 KW DNA-binding; Coiled coil.
 FT DOMAIN 12 100 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 107 798 COILED COIL (POTENTIAL).
 FT DOMAIN 117 120 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 679 682 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 880 883 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 981 989 ARG/LYS-RICH (BASIC).
 FT CONFLICT 46 46 L -> F (IN REF. 2).
 FT CONFLICT 106 106 F -> Y (IN REF. 2).
 FT CONFLICT 153 153 F -> C (IN REF. 2).
 FT CONFLICT 161 161 K -> T (IN REF. 2).
 FT CONFLICT 168 168 E -> D (IN REF. 2).
 FT CONFLICT 216 216 N -> S (IN REF. 2).
 FT CONFLICT 225 226 HG -> FE (IN REF. 2).
 FT CONFLICT 350 350 K -> N (IN REF. 2).
 FT CONFLICT 360 360 E -> D (IN REF. 2).
 FT CONFLICT 400 401 KN -> NY (IN REF. 2).
 FT CONFLICT 406 406 K -> I (IN REF. 2).
 FT CONFLICT 415 415 K -> T (IN REF. 2).
 FT CONFLICT 449 449 E -> D (IN REF. 2).
 FT CONFLICT 483 510 IQLTALTTSEQYKSEVDLXTELENEK -> YSVCHVKW
 FT CONFLICT 516 528 TVLPKRGQRPLSSKRE (IN REF. 2).
 FT CONFLICT 528 528 LPSHCKUSLENK -> YFTLQQAAPPEN (IN REF. 2).
 FT CONFLICT 549 549 N -> I (IN REF. 2).
 FT CONFLICT 560 560 K -> T (IN REF. 2).
 FT CONFLICT 805 805 E -> D (IN REF. 2).
 FT CONFLICT 941 941 P -> S (IN REF. 2).
 SQ SEQUENCE 976 AA; 114069 MW; 8BA81D042AC2696B CRC64;

Alignment Scores:
 Pred. No.: 0.000746 Length: 976
 Score: 136.00 Matches: 203
 Percent Similarity: 35.47% Conservative: 156
 Best Local Similarity: 20.06% Mismatches: 347
 Query Match: 1.92% Indels: 307
 DB: 1 Gaps: 48

US-09-602-362E-22 (1-4115) x SCPI_HUMAN (1-976)

QY 1410 TTTCCAAATAGCCTTGAATCAAGATGAACAAACATTGAAGCAGATCCG----- 1450
 DB 50 PheAlaLysThrAsnLeuSerLysAsnGlyGluAsnIleAspSerAspProAlaLeuGln 69
 QY 1461 -----TGTTCACCGCAATC 1475
 DB 70 LysValAsnPheLeuProValLeuGluGlnValGlyAsnSerAspCysHisTrpGlnGlu 89
 QY 1476 CAACAAAGCAGCTATAGAAATTTCTGGATTCTAGAGTCTCTGAGACTTTTACAGA 1535
 DB 90 GlyLeuLysAsp-----SerAspLeuGluAsnSerGlyLeuLysTrpArg 104
 QY 1536 AGGATGTCTTTACCAAGGCTACACATCAAAAGAAATAGATAAAATGGAATAGAA 1595
 DB 105 -----ValPheSerLysLeuTrpLysGluAlaGluLysIleLysLysTrpLysVal--- 121
 QY 1596 GAGTCCCTAATAAGTGTCTTCTGAGGCTACCTCGGATCAAAATTTCTATCCACTAA 1655
 DB 122 -----SerThrGluAla----- 125
 QY 1656 AGCTAGATTGAAGACATGCAAACTTTTCAAGCGAGCCTCCGGAGGCCATCTGCTCG 1715

Db 126 -----GluLeuArgGln---LysGluSerLysLeuGlnGluAsnArg-LysIleIleGl 142
 QY 1716 AGCTCCACTGAAATCAAAAGTCTGTCCAAATAAGCCCTTGAATGAATAAATAAACA 1775
 Db 142 uAla-----GlnArgLysAlaIleGlnLeuGlnPheGlyAsnGluLys-ValSerL 160
 QY 1776 TGGAGCAGATGAGAACTCCCATCAGATCCAAACAAAGAGACTATAAGAAATTCGGGA 1835
 Db 160 eutLysLeuGluGluGlyIleGlnGlu-----AsnLysAspLeuLysGluAsnAsn 178
 QY 1836 TACTGGAGTCTCTGAGACTGTTTACAGAAAGATTTGTCTTACCAAGGCTCGCTCAAA 1895
 Db 178 laThrArgHisLeuCysAsnLeuLeuLysGluThrCys-----AlaArgSerA 194
 QY 1896 AGAATAGATAAAATTAATGGAATAATAGAGGTCCTCTTAAAGTGTCTTCTGAGGCT 1955
 Db 194 laGluLysThrLysLysTrpGluTrpGluArgGluGluThrArgGlnValTrp----- 211
 QY 1956 AACTGGGAATGAAAGTCTTATCCAACTAAAGCTAGAAATTCATGATGATGCAAACTCAA 2015
 Db 212 -----MetAspLeuAsnAsnIleGluLysMetIleThrAlaH 225
 QY 2016 AG-----CAGGCTCCCGAGAGCCATCTGCCTTCAGCCTCCATGAAATCAAAAG 2066
 Db 225 isGlyGluLeuArgValGlnAlaGluAsnSerArgLeuGluMethHisPheLysLeuLysG 245
 QY 2067 TCTGTCAAATAAACCTTGAATTAAGATGAACAAACATTCAGGCAGATGAGAACTCC 2126
 Db 245 luAspTrpGluLys-----IleGlnHisLeuGluGlnGluT 257
 QY 2127 ATCAATCCAAACAAAGACTATGAGAAAGTTCTTGGATTCTGGAGTCTCTGTAGACTG 2186
 Db 257 yrLysLysGluIleAsnAspLysGluLysGlnValSerLeuLeu----- 271
 QY 2187 TTTCCAGAGGATGGTGTTCACCGCTACATCAAAAGAAATAGATAAA-----T 2240
 Db 272 -----LeuIleGlnIleThrGluLysGluAsnLysMetLysAspLeuT 286
 QY 2241 AATGAAATAGAGAGTCTCTGATAATGAGGTTTTCTGAAGGCTCCCTGAGAAATGAAG 2300
 Db 286 hrPheLeuLeuGluGluSer-----ArgAspLysValAsnGlnLeuGluLysT 303
 QY 2301 TTCTATTCACATAAGCCCTTGAATTCATGCGATGCAAACTTT---CAAAGCGGCTCC 2357
 Db 303 hr-----LysLeuGlnSer-GluAsnLeuLysGlnSerIleGluLys 316
 QY 2358 GAGAGCCATCTGCTCGAGCTGCATTGAAATGAAAGTCTGTCCAAATAAACCTTGA 2417
 Db 317 GlnHisLeuThrLysGluLeuGluAspIleLysValSerLeuGlnArgSerValSer 336
 QY 2418 ATTAAGATGAACAAACATTGAGAC----- 2441
 Db 337 ThrGlnLysAlaLeuGluGluAspLeuGlnIleAlaThrLysThrIleCysGlnLeuThr 356
 QY 2442 -----AGATCAGATTCCCTTCCAGATCAAAACAAAGAGGTTGAGAAATTTCTCG 2492
 Db 357 GluGluLysGluThrGlnMetGluGluSerAsnLysAlaArgAlaAlaHisSerPheVal 376
 QY 2493 GATTCCTGAAGTCTCCCTGGAC---TGTTCACAGAAAGGATGGTGTATCCCGGCTACAC 2549
 Db 377 ValThrGluPheGluThrValCysSerLeuGluGlu-----LeuLeuArgThrGlu 394
 QY 2550 ACAAAGAAGAAATAGTAGTGGAAATTAGAAGATTCAGCTAGCTATCAAAATCTTG 2609
 Db 395 GlnGlnArgLeuGluLysAsnGluAspGln-----LeuLysIleLeu 408
 QY 2610 GTACAGTTATCTGTGAAAGCGCAAGGAACTTCAAAAGACACTGTGAACGCTACAG 2669
 Db 409 ThrMetGluLeuGlnLysLysSerSerGluLeuGluGluMetThrLysLeuThrAsnAsn 428
 QY 2670 AAA-----ATGAACAAATGAAGAGTTTG----- 2696
 Db 111

Db 429 LysGluValGluLeuGluLeuLysLysValLeuGlyGluLysGluThrLeuLeuTyr 448
QY 2696 ----- 2696
Db 449 GluAsnLysGlnPheGluLysLeuAlaGluGluLeuLysGlyThrGluGlnGluLeuLe 468
QY 2697 -----TGTACTGAAAAAAGAACTGTCAGAACAAAGAAATAAATACACA----- 2738
Db 469 GlyLeuLeuGlnAlaArgGluLysGluValHisAspLeuGluLeuGlnLeuThrAlaLe 488
QY 2739 -----GTAGAGAAC 2747
Db 489 ThrThrSerGluGlnTyrTyrSerLysGluValLysAspLeuLysThrGluLeuGluAsn 508
QY 2748 CAAAGTTA---ATGGGAACAGAGTCTGAGTGTAGATTGACTTAACCAAGAAAGA 2804
Db 509 GluLysLeuLysAsnThrGluLeuThrSerHisCysAsnLysLeuSerLeuGluAsnLys 528
QY 2805 GAA-----GAGAGAAATCGCGATATA-----TTAAAGAAAAAATTAGGAGAAAT 2849
Db 529 GluLeuThrGlnGluThrSerAspMetThrLeuGluLeuLysAsnGlnGlnGluAspLe 548
QY 2850 AGAAGATCGAGAGCAGATAGAAAGATTAGAGTGAACACAACTTGAAGCTCTCA 2909
Db 549 AsnAsnAsnLysLysGlnGluGluArgMetLeuLys-----GlnGluGluAsnGln 566
QY 2910 GATACAGATAGAAATTCGAAGTGTAGAAAGTAATTTGAATAGCTTCTCAACTCATCAA 2969
Db 567 GluThrGluThrGlnLeuArg-----AsnGlu 575
QY 2970 ATGAAATATTCTTACATGAAATTCGATGTGAAAGAAAGAAATTCGATGCAAACTGGA 3029
Db 576 LeuGluTyrValArgGluGluLeuLysGlnLysArgAspGluValLysCysLysLeuAsp 595
QY 3030 ATAGCCACACTAAACACCAAT-----CCAGGAAAGGAAATTAATC 3071
Db 596 LysSerGluGluAsnCysAsnLeuArgLysGlnValGluAsnLysAsnLysTyrLe 615
QY 3072 TTTGAGGACTTAAGATTTTAAAGAAAGATGCTGAATCTTAGATGACCCCTAAACTGAAAA 3131
Db 616 ---GluGluLeuGlnGlnGluAsnLysAlaLeuLysLys-----Lys 628
QY 3132 GGAATCATTAAGAGGATCTCAATATGTGGCAGCTAAAGTT----- 3176
Db 629 GlyThrAlaGluSerLysGlnLeuAsnValTyrGluLeLysValAsnLysLeuGluLeu 648
QY 3177 ---CTGATAGCTGAGAAACATGCTCAATCTAAAT-----GAGGAAAAA 3218
Db 649 GluLeuGluSerAlaLysGlnLysPheGlyGluLeuThrAspThrTyrGlnLysGluLe 668
QY 3219 CAGACAAAGAAACTAGAGGAGAAATTTGAATACACCATCTGACTGGCTTCGTGTACA 3278
Db 669 GluAspLysLysIleSerGluGluAsnLeuLeuGluValGluLysAlaLysValLe 688
QY 3279 AGCCATGATCAATTGTGACATCAAGAAAGTA---AGAACCTGCTTCCAC---ATTGAG 3332
Db 689 AlaAspGluAlaValLysLeuGluLeuLysAspLysArgCysGlnHisLysIleAla 708
QY 3333 GAGATGCTGTTTGCACAAAGAAATGAATGTTGATGTGAGAGTACCATATTAACAATGAGTG 3392
Db 709 GluMetValAlaLeuMetGluLys----- 716
QY 3393 CTCCATCACAC-----TTTCTG 3410
Db 717 ---HisLysHisGlnTyrAspLysIleLeuGluArgAspSerGluLeuGlyLeuTyr 735
QY 3411 AGCTCAAG-----GAAATCCAAAC 3431
Db 736 LysSerLysGluGlnGlnSerSerLeuArgAlaSerLeuGluLeuGluLeuSerAsn 755
QY 3432 CTAAGAA-----ATTATCTCAATTAGCAGGAGATGCTTACAGAGAAATACATGTTTTCAG 3485
Db 756 LeuLysAlaGluLeuSerValLysGlnLeuGluArgGluGluLysGlu 775

QY 3486 AACATCACAAGAGACACAGTGAACACAGTGTCAATGAAGGAGGACACACATGTATCAA 3545
Db 776 LysLeuArgGluAlaLysGluAsnThrAlaThrLeuLysGluLysAspLysLys 795
QY 3546 AACGACAAGATATGACAAACACACCTGAA---CAGCAGAGTCTCTAGTCAGAAATATT 3602
Db 796 ThrGlnThrPheLeuLeuGluThrProGluIleTyrTrpLysLeuAspSerLysAlaVal 815
QY 3603 TCAACTACAGCAAAATTTGCTGCTTCACAGCAATTTAGTCACATAGAAAG----- 3656
Db 816 ProSerGlnThrValSerArgAsnPheThrSerValAspHisGlyIleSerLysAspLys 835
QY 3657 -----CTGACACAAACACAGATAACAATGA-----TAT 3685
Db 836 ArgAspTyrLeuThrThrSerAlaLysAsnThrLeuSerThrProLeuProLysAlaTyr 855
QY 3686 TCATTTCTTTCAGAGGAAATGCAACACATCTCTCTAAAGAGAAAGAGAGATATTAAAT 3745
Db 856 ThrVal-----LysThrProThrLysProLysLeuGlnArgGluAsnLeu--- 871
QY 3746 ACATACCATTTAAACACCGTATATTCATATATGAAAGAGAGAGAGAGAGAGAGAGAG 3793
Db 872 AsnIleProIleGlu-----SerLysLysLysArgLysMetAlaPheGluPhe 888
QY 3794 -----AACAGAAACTCATGAGAACAA 3814
Db 889 AspIleAsnSerAspSerSerGluThrThrAspLeuLeuSerMetValSerGluGluGlu 908
QY 3815 GCAGTAGAACTTCTTGGAGAAACACAGACAGACAGCTTTACTCAC-----AC 3862
Db 909 ThrLeuLysThr-LeuTyrArgAsnAsnProAlaSerHisLeuCysValLysTh 928
QY 3863 TCATGCTAGAGCCAGCTCTAGCATCACC 3890
Db 928 rProLysLysAlaProSerSerLeuThr 937
RESULT 5
NUSM CAEEL
ID NUSM CAEEL STANDARD; PRT; 527 AA.
AC P24856;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 5 (EC 1.6.5.3).
GN ND5 OR MTCE.35.
OS Caenorhabditis elegans.
OG Mitochondrion.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=92201635; PubMed=1551572;
RA Okimoto R., Macfarlane J.L., Clary D.O., Wolstenholme D.R.;
RT "The mitochondrial genomes of two nematodes, Caenorhabditis elegans
and Ascaris suum.";
RL Genetics 130:471-498 (1992).
RN [2]
RP SEQUENCE OF 1-120 FROM N.A., AND VARIANT LEU-94.
RC STRAIN=AB1 AB2, Bristol N2, CB4852, CB4853, CB4854, CB4855, CB4856,
RC CB4857, CB4858, KR314, PB303, PB306, RW7000, and TR403;
RX MEDLINE=22531585; PubMed=12644560;
RA Denver D.R., Morris K., Thomas W.K.;
RT "Phylogenetics in Caenorhabditis elegans: an analysis of divergence
and outcrossing.";
RL Mol. Biol. Evol. 20:393-400 (2003).
CC -I- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

QY 2424 TCCTAATCCAGGCTTTATTGGACAGACTTTTCATTTCAATCGAGGCTCGAAGCAGAT 2365
 Db 449 heLeuLeuLeuMetLeuLeuMetLeuLeuMetLeuLeuVal-----AspT 467
 QY 2364 GCCTCTCGGAGGCTGCTTTGAAGTTTGCATGCCATCAATCAAGCTTTAGTGAAT 2305
 Db 467 yrLeuAlaLeuAsnSerLeuLeuLeuMetLeuLeuMetLeuLeuVal----- 487
 QY 2304 AGAATCTTCATTCACGGA-----GCCTTCAGAAACCTCAT 2266
 Db 487 snAsnLeuAsnSerLeuLeuLeuMetLeuLeuMetLeuLeuVal----- 507
 QY 2265 ATCAGAGACTCTCTCTATTT-----CCATTTATTTATCTATTTTGTATGT 2218
 Db 507 yrLeuLeuSerLeuLeuLeuMetLeuLeuMetLeuLeuVal----- 527
 QY 2217 GT 2216
 Db 527 ys 527
 RESULT 6
 REST_CHICK
 ID REST_CHICK STANDARD; PRT: 1433 AA.
 AC 042184; 042228; 057563; 057564;
 DT 15-JUL-1998 (rel. 36, Created)
 DT 15-JUL-1998 (rel. 36, Last sequence update)
 DT 10-OCT-2003 (rel. 42, Last annotation update)
 DE Restin (Cytoplasmic linker protein-170) (CLIP-170).
 GN RSN.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98137792; PubMed=9469933;
 RA Griparic L., Volosky J.M., Keller T.C. III;
 RT "Cloning and expression of chicken CLIP-170 and restin isoforms.";
 RL Gene 206:195-208(1998).
 CC [2]
 CC SEQUENCE OF 17-1139 FROM N.A. (ISOFORMS 3 AND 4).
 CC TISSUE=Pectoralis muscle;
 CC MEDLINE=99002898; PubMed=9784600;
 CC Griparic L., Keller T.C. III;
 CC "Identification and expression of two novel CLIP-170/Restin isoforms
 CC expressed predominantly in muscle.";
 CC Biochim. Biophys. Acta 1405:35-46(1998).
 CC -!- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
 CC THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
 CC CYTOSKELETON (BY SIMILARITY).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=042184-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=042184-2; Sequence=VSP_000761;
 CC Name=3; Synonyms=CLIP-170(11);
 CC IsoId=042184-3; Sequence=VSP_000762, VSP_000763;
 CC Name=4; Synonyms=CLIP-170(11+35);
 CC IsoId=042184-4; Sequence=VSP_000764;
 CC -!- SIMILARITY: Contains 2 CAP-Gly domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL curation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AF014012; AAC60344.1; -
 DR EMBL; AF020764; AAC60345.1; -
 DR EMBL; AF045650; AAC03547.1; -
 DR EMBL; AF045651; AAC03548.1; -
 DR InterPro; IPR000938; CAP-Gly.
 DR InterPro; IPR001878; Znf_CCHC.
 DR Pfam; PF01302; CAP_GLY; 2.
 DR SMART; SM00343; Znf_C2HC; 1.
 DR PROSITE; PS00845; CAP_GLY_1; 2.
 DR PROSITE; PS02445; CAP_GLY_2; 2.
 KW Cytoskeleton; Microtubule; Coiled coil; Repeat; Alternative splicing.
 FT DOMAIN 79 121
 FT CAP-GLY 1.
 FT SER-RICH.
 FT CAP-GLY 2.
 FT DOMAIN 235 277
 FT SER-RICH.
 FT DOMAIN 305 332
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 351 1353
 FT CCHC-BOX.
 FT DOMAIN 1414 1427
 FT Missing (in isoform 2).
 FT VARSPLIC 458 492
 FT /FTID=VSP_000761.
 FT TOTLEHARKEQLSEKLLFEKTKADKLORELEDTTR -> RK
 FT ROISDPEN (in isoform 3).
 FT /FTID=VSP_000762.
 FT S -> GSSKVS (in isoform 3).
 FT VARSPLIC 803 803
 FT /FTID=VSP_000763.
 FT T -> RKQISDPEN (in isoform 4).
 FT VARSPLIC 458 458
 FT /FTID=VSP_000764.
 FT CONFLICT 309 309
 FT K -> R (IN REF. 2; AAC03547).
 FT CONFLICT 440 440
 FT E -> V (IN REF. 2; AAC03548).
 SQ SEQUENCE 1433 AA; 161026 MW; 5631CE8683498E23 CRC64;
 Alignment Scores:
 Pred. No.: 0.00246 Length: 1433
 Score: 130.50 Matches: 231
 Percent Similarity: 36.01% Conservative: 191
 Best Local Similarity: 19.71% Mismatches: 386
 Query Match: 1.84% Indels: 365
 DB: 1 Gaps: 54
 US-09-602-362e-22 (1-4115) x REST_CHICK (1-1433)
 QY 892 ACATCTGGAAGTTCGAGCTCAGCAGAGAAACACACTAGGAATACGATCCTGCAAAA 951
 Db 340 ThrSerSerArgTyrAlaArgLysIleSerGlyThr-----ThrAlaLeuGluGlu 356
 QY 952 ACATCTGGAATTTACGGCCA-----GCAAGAGACCTAGAGAT 996
 Db 357 AlaLeuLysGluLysGlnGlnHisIleGluGlnLeuLeuAlaGluArgAspLeuGluArg 376
 QY 997 CGCAGCGAGAAAGAGAGACA--CACCTAGGAAATATGACTCCGAAAGAAACACATG 1053
 Db 377 AlaGluValAlaLysAlaThrSerHisValGlyGluIleGluGlnGluLeuAlaLeuVal 396
 QY 1054 AGAAATTACGTGGCAGAGAAAGAGACCTAGGAGACGATGGAGAGAAAGAACCCCT 1113
 Db 397 ArgAsp-----GlyHisAspArgHisVal 404
 QY 1114 GTAAGCTGGATCGCTGCAAGAGTAACATCTAATAACTAAAGTTTGTAAAGGAATC 1173
 Db 405 LeuGluMetGluAlaLysMetAspGlnLeu-----ArgAlaMet 417
 QY 1174 TAAGATGTTGCATGTCACCAAGAAAGATCATCTACAAAGCAAGTGCATGATCAGAGTCC 1233
 Db 418 ValGluAlaAlaAspArgGluLysValGluLeuLeuAsnGlnLeuGluGluLysArg 437
 QY 1234 -----CATCATCCAAACAGGAGATGAAGATATTCTTGAATTCGGGTCTCTT 1287
 Db 438 LysValGluAspLeuGlnPheArgValGluGluGlu-----SerIleThrLysGlyAspLeu 456
 QY 1288 TGAAGTTCT-----GCAAGATTCAAGTGGTATACCTGAGTCT--ATATA 1329
 Db 457 GluThrGlnThrLysLeuGluHisAlaArgIleLys-GluLeuGluGlnSerLeuLeuPh 476

Db 154 SerTyrTyrPheLeuSerLeuSerPhePhe-----CysTrpLeuSerSer 168
 QY 3277 -----CTACAGGAGCAGTCTAGGATGCTGT 3251
 Db 169 LeuMetLeuLeuAlaSerPheThrLysSerAlaGlnPhePropheSerGlyTrp--- 187
 QY 3250 ATTCAATTTCTGCTT----- 3236
 Db 188 -----LeuProLysAlaMetSerAlaProThrProLysSerSerLeuValHisSer 204
 QY 3235 -----CTAGTTTCTTCTGCTGCTTTTCTTCTCAATTTAGATGAG 3197
 Db 205 SerThrLeuValThrAlaGlyLeuValLeuIleMetAsnPheSerGluMet----- 221
 QY 3196 CATTTCTGCTATCAGAACTTTAGCTGCCACATATTGATGATCCCTTTTAGTAATG 3137
 Db 222 -----IleLeuAsnLysAspValIleMet 229
 QY 3136 ATTCCTTTTTCAGTTTTCAGGTCATCTAGTTTCAGCATCTTTTCTTAAATCTTAACTTC 3077
 Db 230 Ile--IleMetValValGlyValPheThrMetPhePheSerSerMetAlaAlaLeuValG 249
 QY 3076 TCAAGATTTA-----TTTTCTCTTTCTCTGGA 3050
 Db 249 IuGluAspLeuLysLysValValAlaLeuSerThrLeuSerGlnMetGlyPheSerMetL 269
 QY 3049 TTGGTGTTCAGTGTGCTATTCAGTTTTCAGTGTGCTATTCCTTTTCAACATCAATT- 2991
 Db 269 eutThrValGlyIleGlyLeuSerPheValSerPheIleHisLeuSerHisAlaLeuP 289
 QY 2990 -----TCATGTAGAATAATTTTCATTCATG----- 2964
 Db 289 heLysSerCysLeu-----PheMetGlnValGlyTyrLeuIleHisCysSerL 305
 QY 2963 -----AGTTGAGAACTTATCAAAAT-----TACTTTCTCAACC 2930
 Db 305 euGlyGlnAspGlyArgSerLysLeuGlyAsnValProTyrPheIleGlnL 325
 QY 2929 TTCAATTCCTTCTCTGATCTCTGATCTGAGACCTTTCAAGTTCTGT----- 2899
 Db 325 euGlnLeuValThrLeu-----PheCysLeuCysGlyLeuValPheSerSerG 342
 QY 2888 -----TTCACTTCTAACTCTTTCTA----- 2869
 Db 342 lyAlaValSerLysAspTyrIleLeuGluPhePhePheSerAsnPheMetValValP 362
 QY 2868 --TGCTGCTCTCGATCTCTTAATTTCTCTCAATTTTCTTAAATATATCGCATTTCT 2811
 Db 362 heAlaCysMetPhePhePheSerValPheLeuThrPheGlyTyrSerTyrArgLeuTrpL 382
 QY 2810 CTCCTCTCTTT--TCTTGTTTAACTCAATCTACACGACCTCTGTTCCCATTTAA 2754
 Db 382 ysGlyPhePheMetSerPheSerArgProValPheCysPheSerSerSerValValMetA 402
 QY 2753 CTTTGGTGTCTCTACTGCTATTTCTTTCTGCTCTGACAGTCTTTTTCAGTACACAA 2694
 Db 402 snPheLeuSerLeu-----LeuLeuValLeuPhe-----S 412
 QY 2693 AACTCTTTTTCATT-----GTTTCATTTCTGTCAGTGTTCACAGTCTTTTTCAGG 2640
 Db 412 exIlePhePheIleTrpTrpMetAsnPheAsnMetLeuCysMetProCysLeuPheLeuT 432
 QY 2639 TTCCCTTGTCTTCAACAGATCACTGTACCAAGATTTTGTAGGCTAGTGAATCTTCTA 2580
 Db 432 yrValAspPhePheVal----- 437
 QY 2579 ATTTTCCACTATTTATCTTTTC-----TTTTGTGTGTA----- 2545
 Db 438 -----ProLeuPhePheValMetIleMetValValGlyPheLeuCysValLysLeuL 456
 QY 2544 -----GCCTGGGTACACA 2532
 Db 456 euLeuLysGluPheValTyrLysPheLeuValAspPhePheAlaLysGlyTrpValTyr- 475

QY 2531 CCATCCTTCTGTGAAACAGTCCACGGAGACTTCAGAAATCCAGAAATTTTCTCAACCTTCT 2472
 Db 476 -----GlyLeuLysAsnTyrLysPhePheAspLeuPheL 487
 QY 2471 TTGTTTTGATTCGAAGGAAATCTGATCTCTCAATGTTGTTTCATTTTAATTTCCAAG 2412
 Db 487 eu-----GlyGlyIleAsnSerLeuGlyValThrPhePheSerPheThrG 502
 QY 2411 GTTTATTGGAACAGACTTT 2392
 Db 502 lyPheTrpSerAsnSerTyr 508
 RESULT 8
 SBCC_CLOAB STANDARD; PRT; 1163 AA.
 ID SBCC_CLOAB
 AC Q97FK1
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nuclease sbcCD subunit C.
 GN SBCC OR CAC2736.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 CX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RA MEDLINE=21359325; PubMed=11466286;
 RX Noelling J., Berton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Ratsov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.,
 RA "Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum";
 RL J. Bacteriol. 183:4823-4838(2001).
 CC -I- FUNCTION: SbcCD cleaves DNA hairpin structures. These structures
 CC can inhibit DNA replication and are intermediates in certain DNA
 CC strand exonuclease reactions. The complex acts as a 3'->5' double
 CC single-strand endonuclease activity (by similarity).
 CC -I- SUBUNIT: Heterodimer of sbcC and sbcD (by similarity).
 CC -I- SIMILARITY: Belongs to the SMC family. SbcC subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE007771; AAK80682.1; --
 DR PIR; G97236; G97236.
 DR InterPro; IPR003439; ABC transporter.
 DR HydroLase; Nuclease; Exonuclease; Endonuclease; DNA replication;
 KW DNA recombination; ATP-binding; Coiled coil; Complete proteome.
 FT NP BIND 35 42 ATP (POTENTIAL).
 FT DOMAIN 197 415 COILED COIL (POTENTIAL).
 FT DOMAIN 446 1003 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1163 AA; 135507 MW; C85F0BD215D7A92 CRC64;
 Alignment Scores:
 Pred. No.: 0.00264 Length: 1163
 Score: 130.00 Matches: 197
 Percent Similarity: 39.48% Conservative: 150
 Best Local Similarity: 22.41% Mismatches: 278
 Query Match: 1.83% Indels: 255
 DB: 46
 US-09-602-362E-22 (1-4115) x SBCC_CLOAB (1-1163)


```

Db      856  erAspGluillellystYrHisSerAsnLeuSerLeuLysAspArgLysValAsnA 876
QY      3439  TTATCTCAATTAGCAGGAGTGTCTAGAGAAATACATGTTTCAGAACATCACAAAGA 3498
Db      876  spileAspLysLeuAsnLysleileuMetGlu---GluLysPheGluAsnlelGluLys- 894
QY      3499  GACACGCGTGAACAGTGTCAATGAAGGAGGACGACACATCATATCAAAACGACAGATAA 3558
Db      895  --AlaLysGluAsnTyrLeuAsnAspLysGluileAsnLeuLysSerAsp-----V 912
QY      3559  TGGAAACAAACCTGAAACGACGAGTCTCTAGTCA-----GAAATATTTCACTA 3609
Db      912  alGluLysTyrLysAsnGluLeuSerLysValAsnGlyAlaValGluValleuSerLysL 932
QY      3610  CAAGCAAAATTTGGTCTCACAGCAATAGTATGACATAGAAAGCTGACA----- 3662
Db      932  ysLeuLysAsn-----ArgLysLeuThrGluLul 942
QY      3663  -----ACAAAGCAGATACAAATGATATCAT-----TTCTTGAGAG 3699
Db      942  ysTrpIleGluillelgnAsnAsnArgValGluLysAlaSerLysAlaLysAlaLeuG 962
QY      3700  GAAATGCAACATCTCTAAAGAGAAAGAGAGAGATATTAATCAATACCATTTAA 3759
Db      962  luArgSerileLysleuGluGluValLysAsn-----IleGluileLysleuL 979
QY      3760  AAACCGTATATTCATATCAATGAAAGAGAAAGAGAAACAGAA 3800
Db      979  ysGluLeuGlyLysLeuLysLysThrLysGlnGluLeuGlu 992

RESULT 9
GOA4 HUMAN
ID GOA4 HUMAN STANDARD; PRT: 2230 AA.
AC Q13439; Q13270; Q13654; Q14436;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Golgi autoantigen, golgin subfamily A member 4 (Trans-Golgi p230) (256
DE KDA golgin) (Golgin-245) (72.1 protein).
GN GOLGA4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Butheria; Primates; Catarrhini; Hominidae; Homo.
OY NCBI_TaxID=9606;
RN [1] SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RP MEDLINE=96215236; PubMed=8626529;
RA Erlich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.;
RT "Molecular characterization of trans-Golgi p230: a human peripheral
RT membrane protein encoded by a gene on chromosome 6p12-22 contains
RT extensive coiled-coil alpha-helical domains and a granin motif.";
RL J. Biol. Chem. 271:8328-8337(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Seelig H.P.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 131-2230 FROM N.A.
RC TISSUE=Placenta; PubMed=8537393;
RX MEDLINE=96125112; PubMed=8537393;
RA Fritzler M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.L.;
RT "Molecular characterization of golgin-245, a novel Golgi complex
RT protein containing a granin signature.";
RL J. Biol. Chem. 270:31262-31268(1995).
RN [4]
RP SEQUENCE OF 524-672 FROM N.A.
RC TISSUE=Gastric fundus;
RA Balague C.;
RL Thesis (1994), Instituto municipal de investigacion medica, Spain.
CC -!- FUNCTION: May play a role in vesicular transport from the trans-
CC Golgi.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic, peripheral membrane protein
CC associated with the trans-Golgi network.

```

```

CC      -!- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=4;
CC      Comment=Additional isoforms seem to exist;
CC      Name=1;
CC      IsoId=Q13439-1; Sequence=Displayed;
CC      Name=2;
CC      IsoId=Q13439-2; Sequence=VSP_004272; VSP_004273;
CC      Name=3;
CC      IsoId=Q13439-3; Sequence=VSP_004274;
CC      Name=4;
CC      IsoId=Q13439-4; Sequence=VSP_004275;
CC      -!- DOMAIN: Extended rod-like protein with coiled-coil domains.
CC      -!- DISEASE: Antibodies against GOLGA4 are present in sera from
CC      patients with Sjogren's syndrome (SS) [MIM:270150]. Sera from
CC      patients with SS often contain antibodies that react with normal
CC      components of the Golgi complex.
CC      -!- DISEASE: Antibodies against GOLGA4 are found in sera from
CC      hepatitis B patients.
CC      -!- SIMILARITY: Belongs to the golgin family.
CC      -!- SIMILARITY: Contains 1 GRIP domain.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      EMBL; U41740; AAC50434.1; -
CC      EMBL; X82834; CAA58041.1; -
CC      EMBL; U31906; AAC51791.1; -
CC      EMBL; X76942; CAA54261.1; -
CC      Genew; HGNC:4427; GOLGA4.
CC      MIM; 602509; -
CC      GO; GO:0005802; C:Golgi trans face; TAS.
CC      GO; GO:0016192; P:vesicle-mediated transport; TAS.
CC      InterPro; IPR000237; GRIP_domain.
CC      Pfam; PF01465; GRIP; 1.
CC      PROSITE; PS50913; GRIP; 1.
CC      Antigen; Golgi stack; Coiled coil; Alternative splicing.
CC      FT DOMAIN 133 2185 COILED COIL (POTENTIAL).
CC      FT DOMAIN 252 2096 GLU-RICH.
CC      FT DOMAIN 2168 2215 GRIP.
CC      FT VARSPLIC 2154 2185
CC      TPYKGNLYHTDVSFLFGEPEYLRKVLFEY -> HLTKV
CC      AICTRMSHSLNPLNLSICKCFLSI (in isoform
CC      2).
CC      /FTID=VSP_004272.
CC      Missing (in isoform 2).
CC      /FTID=VSP_004273.
CC      Missing (in isoform 3).
CC      /FTID=VSP_004274.
CC      FTSPRSRGIF -> SWLRSSS (in isoform 4).
CC      /FTID=VSP_004275.
CC      R -> K (IN REF. 3).
CC      Y -> H (IN REF. 3).
CC      T -> A (IN REF. 3).
CC      K -> E (IN REF. 3).
CC      T -> A (IN REF. 3).
CC      K -> E (IN REF. 3).
CC      K -> N (IN REF. 3).
CC      K -> N (IN REF. 3).
CC      SQ SSSEQUENCE 2230 AA; 261139 MW; 3BB733DB1EA86134 CRC64;

Alignment Scores:
Pred. No.: 0.00291 Length: 2230
Score: 130.00 Matches: 234
Percent Similarity: 34.68% Conservative: 179
Best Local Similarity: 19.65% Mismatches: 382
Query Match: 1.83% Indels: 396
DB: 1 Gaps: 57

US-09-602-362E-22 (1-4115) x GOA4_HUMAN (1-2230)

```

| | | | |
|----|------|--|------|
| QY | 820 | CTGTTGAAAAACACCTGATGAGCGTTCATCTTGGTGGAGAACATCTGACAAATTCAT | 879 |
| Db | 608 | MetValGluLysHisLysThrGluLeuGlu | 617 |
| QY | 880 | GTTTGAGAAAGACATCTCGAAAGATTCGAACGTCAGCAGAAAAACACCTAGAAATTCAG | 939 |
| Db | 618 | -----SerLeuLysHisGlnGlnAspAlaLeuThrThrGluLysLeuGln | 632 |
| QY | 940 | ATCTGTGCAAAAAACATCTGAGAAATTTACGGCGCAGCAAGGAAAGACCTAGAGATCGC | 999 |
| Db | 633 | ValLeuLysGlnGln-----TyrGlnThrGluMetGluLysLeuArgGluLys | 648 |
| QY | 1000 | AGGAGAAAAAGAGACACACTAGGAAA-----TATGAGTCCCGAAA | 1044 |
| Db | 649 | CysGluGlnGluLysGluThrLeuLeuLysAspLysGluLeilepneGlnAlaHisile | 668 |
| QY | 1045 | GAACACCTGAGAAATTCGTGGCGAGAAAAAGACCTAGGAGACGCATCGGAGAAA | 1104 |
| Db | 669 | GluGluMetAsnGluLysThrLeuGluLysLeuAspValLysGlnThr----- | 684 |
| QY | 1105 | AGAAACCTGTAAAGATCGATCGCTGCGAAGGTAACTCTAATAAACTAAAGTTTCAAA | 1164 |
| Db | 685 | -----GluLeuGluSerLeuSerGluLeu-----SerGluValLeuLys | 698 |
| QY | 1165 | AAGAAA-----ATCTAAGATGTTCATGTCCACAAAAG | 1197 |
| Db | 699 | AlaArgHisLysLeuGluGluLeuSerValLeuLysAspGlnThrAspLysMetLys | 718 |
| QY | 1198 | AATCATCTACAAAGCAAGTGCATGATCAGAGTCCCATCAGATCCAAACAGAGGAAGAT | 1257 |
| Db | 719 | GlnGluLeuGluAlaLysMetAspGluGlnLysAsnHisHis-----GlnGlnGlnValAsp | 737 |
| QY | 1258 | GAAGAAATATCTTTGATCTCGGTCTCTTTG-----AAGTCTCGAAGATTC | 1305 |
| Db | 738 | Ser--IleIleLysGluHisGluValSerIleGlnArgThrGluLysAlaLeuLysAspG | 757 |
| QY | 1306 | AAGTGTATACCTGAGTCTATATATAAAAGTAATAGATAAATAAGAAAG----- | 1355 |
| Db | 757 | IleAsnGlnLeuGluLeuLeuLysGluArgAspLysHisLeuLysGluHisGlnAla | 777 |
| QY | 1356 | -----TAGAAGACCT | 1365 |
| Db | 777 | laHisValGluAsnLeuGluAlaAspIleLysArgSerGluGlyGluLeuGlnGlnAla | 797 |
| QY | 1366 | CCTAAAAAGCCATCTGCCTTCAAG----- | 1388 |
| Db | 797 | erAlaLysLeuAspValPheGlnSerTyrGlnSerAlaThrHisGluGlnThrLysAla | 817 |
| QY | 1389 | -----CCGCCATTGAACCAAACTCTTCCAAATAAGCCT----- | 1424 |
| Db | 817 | YrGluGluGlnLeuAlaGluLeuGlnLysLeuLeuAspLeuGluThrGluArgIle | 837 |
| QY | 1425 | --TTGAATGAAGAAATGAACAAACATTGAAGCA-----GATCCGTGT----- | 1463 |
| Db | 837 | euLeuThrLysGlnValAlaGluValGluAlaGlnLysLysAspValCysThrGluLeu | 857 |
| QY | 1464 | --TCCACCGAATCCCAACAAAGACCTATAGAAATTCCT----- | 1502 |
| Db | 857 | spAlaHisLysIleGlnValGlnAspLeuMetGlnGlnLeuGluLysGlnAsnSerGlu | 877 |
| QY | 1503 | -----GGGATTCCTAGAGTCTCTGAGACTGTTTACAGAGGATGTGTTTACCAAGCCTA | 1557 |
| Db | 877 | etGluGlnLysValLysSerLeuThrGlnValTyrGluSer-----LysLeuGluAspCly | 896 |
| QY | 1558 | CACATCAAAAGATAGATAAAAAATCGAAAAATAGAA-----GAGTCCTCTAATAAGTGGT | 1614 |
| Db | 896 | snLysGluGlnGluGlnThrLysGlnIleLeuValGluLysGluAsnMetIleLeuGln | 916 |
| QY | 1615 | CTTCTGAGGCTACCTCGGAATGAAATTTCTATTCCACTAAAGCCCTAGATTGAAACAT | 1674 |
| Db | 916 | etArgGluGlyGlnLysLys-----GluIleGluIleLeuThrG | 929 |

| | | | | |
|----|------|--|---|------|
| QY | 1675 | GCAACATTTTC | AAACGGAGCCCTCCGGGAAGCCATCTGCTTCAGAGCTCCACTGAAATCAAA | 1732 |
| DB | 929 | lnLysLeuSerAlaLys | -----GluAspSerIleHis | 939 |
| QY | 1735 | AGTCTGTCCAAAT | TAAGCCCTTGAATCAAAAATGAAAAATCATGGAGCAGATGAGAACTC | 1794 |
| DB | 940 | -----IleLeuAsnGluGluTrp | -----GluAsnGlnGluLysMetGluL | 958 |
| QY | 1795 | CCATCAGAAATCC | AAACAAAGCATTATTAAGAAAAATTTGGGATACTCGAGTCTCTGTAGAC | 1854 |
| DB | 958 | ysValLysGlnLysAlaLysGluMetGlnGluThrLeu | ----- | 970 |
| QY | 1855 | TGTTTTCACAGAAGGAT | TGTTTACCAGGCTCGCTCAAAAAGATAGATAAAATAATG | 1914 |
| DB | 971 | -----LysLysLysLeuLeuAspGlnGluAlaLysLeuLysLys | ----- | 983 |
| QY | 1915 | GAAATATTAGAAGG | TCCCTGTTTAAAGTGGTCTTCTTGAGGCTAACTGGGAATGAAAGTTCT | 1974 |
| DB | 984 | --GluLeuGluAsnThrAlaLeuGluLeuSerGlnLysGlu | -----LysGlnp | 999 |
| QY | 1975 | ATTCCAACTAAAGC | TAGATTGATGATCGCAACTTCAAAGCAGCCCTCCCGGAGCC | 2034 |
| DB | 999 | heAsnAlaLysMetLeuGluMetAla-GlnAlaAsn | ---SerAlaGlyIleSerAspAla | 1017 |
| QY | 2035 | ATCTGCCTTCGAGCCT | CAATTGAAATCAAAAGTGTCTCCAATAAACCTTGGAAATTAAGA | 2094 |
| DB | 1018 | ValSer | -----ArgLeuGlu | 1022 |
| QY | 2095 | ATGACAAACAT | TGAGCAGATGAGAACTCCCATCAATCCAAACAAAGACATATGAGAA | 2154 |
| DB | 1023 | ThrAsnGlnLys-GluGlnIleGluSerLeuThrGluValHisArgArgGluLeuAsnAs | 1042 | |
| QY | 2155 | AGTTCTTGGCAT | CTCGAGTCTCTGTAGACTGTTTCCAGAAGAGATGGTGTACCACAGCC | 2214 |
| DB | 1042 | pValIleSerIleTrpGlu | -----LysLysLeuAsnGlnGlnAl | 1055 |
| QY | 2215 | TACACATCAAAAG | AATAGATAAAATAATCGAAAAATAGAAAGTCTCTGTATATAGAGTT | 2274 |
| DB | 1055 | aGluGluLeuGlnGlu | -----IleHisGluIleGlnLeuGlnGluLys | 1069 |
| QY | 2275 | TTCTGAA | ----GGCTCCCTGAGAAATGAAGTTCATTCCACTAAAGCCTTGAATTGAT | 2328 |
| DB | 1069 | sGluGlnGluValaGluLeuLysGlnLysIleLeu | -----LeuPh | 1083 |
| QY | 2329 | GGCATGAAACTTT | CAAGCAGCCCTCCGAGAGCCATCTGCTTCGAGCCTGCATTGAAA | 2388 |
| DB | 1083 | eGlyCys | ----- | 1085 |
| QY | 2389 | TGAAAGTCTGTTCCAAATAAA | ----CCTTGGAAATTAAGA | 2427 |
| DB | 1086 | -GluLysGluGluMetAsnLysGluLeThrTrpLeuLysGluGluGlyValLysGlnAs | 1105 | |
| QY | 2428 | AACAAATTCAGACAT | CAGATTCCCTTCAGAAATCAAAACAAAGAGTCTCAGAAAAT | 2487 |
| DB | 1105 | pThrThrLeuAsnGluLeuGlnGluGlnLysGlnLysSerAlaHisValAsnSerLe | 1125 | |
| QY | 2488 | TCGTGGGATTC | GAAGTCTCGTGGACTGTGTTTCACAGAAGATGGTGTGTACCCAGCGCTAC | 2547 |
| DB | 1125 | uAla-GlnAspGluThrLysLeuLysAlaHisLeuGlu | -----LysLeuG | 1140 |
| QY | 2548 | ACAAAAAGAAAGG | ATAAATAGTGMAAATTTAGAAGATTCCTAGCTATCAAAATCT | 2607 |
| DB | 1140 | luValAspLeuAsnLysSerLeuLysGluAsn | ----- | 1150 |
| QY | 2608 | TGGTACAGTT | TCATCTGTGAAAGCGCAAGGAATTCAAAAAGACACTGTGAACACGTACA | 2667 |
| DB | 1151 | -----ThrPheLeuGlnGluGlnLeuValGluLeuLysM | 1162 | |
| QY | 2668 | GGAATATGGAACA | ATGAAAGAGTTTGTGTACTGAAAAAGCACTGTGAGACAAAAAG | 2727 |
| DB | 1162 | etLeuAlaGluGluAspLysArgLysVal | -----SerGluLeuThrSerLysLeuL | 1179 |
| QY | 2728 | AATAAATTCAG | TAGAGAACCAAAAGTTA-----AATGGGAACAAGATC | 2772 |

```
Db 1179 yThrThrAspGluGluPheGlnSerLeuLysSerHisGluLysSerAsnLysSerL 1199 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2773 TGCAGTGTAGATGACTTAAACCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2823 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1199 euGluAspLysSerLeuGluPheLysLysLysSerGluGluLeuAlaLeuAspI 1219 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2824 TATTAAAGAAAAA 2837 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1219 leCysCysLysLysThrGluAlaLeuLeuGluAlaLysThrAsnGluLeuLeuLeu 1239 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2838 -----TTAGGAAGAATTAGAGAAATCGAGAGAGAGAGAGAGAGAGAGAGAGAG 2874 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1239 erSerSerLysThrAsnAlaLeuSerArgLysSerHisCysGlnHisArgThrThrL 1259 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2875 GAGTTAGAGT-----GAAACACAACTTG 2898 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1259 ysValLysGluAlaLeuLeuLysThrCysThrValSerGluLeuGluAlaGlnLeuA 1279 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2899 AAAGGCTCTCAGATACAGATAAGAAATTGAAGGTGTAGAAAGTAATTGTAAGTTTC- 2957 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1279 rgGlnLeuThrGluGluGlnAsnThrLeuAsn-----LysSerPheG 1293 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2958 -----TCAACTCATGAATGAATATTCTTACATGNAATTGCATGTTGAAAAGGAAT 3012 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1293 lnGlnAlaThrHis-Gln----- 1298 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3013 TGCCATGCAAACTGGAATAGCCACTAAACACCAATCCAGAGAAAGGAAATAAATCT 3072 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1299 -----LeuGluGluLysGluAsnGln--- 1305 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3073 TTGAGCACTTAAGATTTAAGAAAGAGTGTGAACCTTAGACCTCAAACTGAAAG 3132 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1306 -----LleLysSerMetLysAlaAspLysGluSerLeuValThrGluLys 1320 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3133 GAATCACTACTAATA-----GGGATCTCAATA 3159 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1321 GluAlaLeuGlnLysGluGlyAsnGlnGlnGlnAlaAlaSerGluLysGluSerCys 1340 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3160 TGTGGCGAGCTAAAGTCTCTGATAGCTGAGAAAC---AATGCTCAATTAATGAGGAAA 3216 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1341 lleThrGlnLysLysGluLeuSerGluAsnLysAlaValThrLeuMetLysGlu 1360 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3217 AAC---AGACAAGAAACTAGAGCAGCAAAATTGAATACACCATCCTGACTGGCTCGCT 3273 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1361 GluLeuLysGluLysValGluLysSerSerLeuSerLysGlnLeuThrAspLeuAsn 1380 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3274 GTACAAGCCATGATCAATTGTGACATCAAGAAAGTAAGAACCTGCTTCCACATGAGG 3333 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1381 ValGlnLeuGlnAsnSerLysSerLeuSerGluLysGluAlaAlaLysSerLeuArg 1400 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3334 AGATGCTGTTCCAAAGAAATGAATGTGTGAGAGTAGTACGATATTAACAAAGTGC 3393 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1401 Lys-----GlnTyAsp---GluGluLysCys 1408 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3394 TCCATCACCACCTTCTGAGCTCAAAGGAATCCAAACCTTAAATATCTCAATTAGCA 3453 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1409 GluLeu-----LeuAspGlnVal--- 1414 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3454 GGAGATGTCTAAGAGAAATACATTTGGTTTCAGAACATCACAAAGAGACAGTGAACA 3513 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1415 -----GlnAspLeuSerPheLys-----ValAspThr 1423 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3514 GTGTCATGAAGAGCAACATGTATCAAAACGACAAAGATAATGG---AACAACAC 3570 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1424 LeuSer---LysGluLysLysSerAlaLeuGlnValAspAspTTPSerAsnLysPhe 1442 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3571 CTGAACAGCAGAGTCTCTAGTCAGAAATATTTCAACTACAGCAAAATTTGTGCTTCA 3630 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1443 SerGluTTPLysLysAlaGlnSerArgPheThrGlnHisGlnAsnThrValLysGlu 1462 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3631 CAGCAATTAGTCATGCATAGAAAGCTCACAA-----CAAAAGCATATACA 3678 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

--!- SIMILARITY: Belongs to the kinesin-like protein family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL curation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/send an email to license@sib-sib.ch>).

EMEL; Z15005; CAA78727.1; -.
 PIR: S28261; S28261.
 HSSP: P17119; 3RAR.
 Genew: HGNC:1856; GENPE.

GK; Q02224; -.
MIM; 117143; -.
GO; GO:0005699; C:kinetochore; TAS.
GO; GO:0005634; C:nucleus; TAS.

GO; GO:0000330; P:kinetochore motor activity; TAS.
GO; GO:0000087; P:DNA replication and chromosome cycle; TAS.
GO; GO:0007079; P:mitotic chromosome movement; TAS.
GO; GO:0007080; P:mitotic metaphase plate congression; TAS.

Pfam; PF00225; kinesin; 1. -
PRINTS; PRO0380; KINESINHEAVY.
SMART; SM00129; KISC; 1.
PROSITE; PS00411; KINESIN MOTOR DOMAIN; 1.

PROSRIE; F55006; KINESIN MOTOR DOMAIN2; 1.
Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
Cell cycle; Centromere; Lipoprotein; Prenylation.
DOMAIN 1 335 KINESIN-MOTOR.

| | | | |
|----------|----------|------------|--------------------------|
| DOMAIN | 2472 | 2663 | GLOBULAR (POTENTIAL). |
| NP BIND | 86 | 93 | ATP (BY SIMILARITY). |
| LIPID | 2660 | 2660 | S-farnesyl cysteine. |
| SEQUENCE | 2663 AA: | 3120A7 MW: | CPECI3A8IC8AC8CB8 CFC64. |

```

ment Scores:
No.:      0.003      length:      2663
:         130.00     Matches:     291

```

```

Local Similarity: 19.90%
Match: 1.83%
Mismatch: 485
Indels: 466
Gaps: 57

```

-602-362E-22 (1-4115) x CENE_HUMAN (1-2663)

537 ATCAGAGATAGTGTGCATGCTTTTAGCAAAATGTGACGCTTTTGTGCAGATATAGTGGAG 596

597 TAACGCGAACAATATGCTGTTACTGTGGATTCTCACATTATATAAAATTATGAATAT 656

657 ATACGAAATTCACAAAATCAACAATAACCAATCAGAAGGAACCTGCAGGAAACTGA 716
::: ::: ||| ||| :::: :||| |
756 VaicluargueuArguysGluiIleGinAspLysSerglu----- 768

717 TGAGGCTGCACCTGGCGGAAGACACCTGCACACGCTGAAGCTGGTGGAAAACACCTG 776
769 -----GluLeuHisIleIleThrSerGluLeuAspLysLeuPheSer 782

777 ATGAGGTGCACCCCTGGTGTGAAGAAGACACTTGACAGCCTGAAGAAGCTGGTGTGAATAAACACCT 836
:::| | | | |
:::| | | | |
783 GluValValHisLysGluSerArgValGlnGlyLeuLeuGluGluIleGlyThrLys 802

837 G A G A G C C G C A
 |||::: |||
 803 A S p A s p L e u A l a T h r G l n S e r A s n T y r - I y s S e r T h r A s p G l n G l u P h e G l n A s n P h 822

[illegible]

| | | | |
|------|----|--|------|
| 822 | Db | eLysThrLeuHisMetAspPheGluGlnLysTyrLys-----MetValle | 837 |
| 945 | Qy | GCAAAAAAC-----ATCTGAGAAATTTACGGCCACGACAAAGAA----- | 984 |
| 837 | Db | uGluGluAsnGluArgMetAsnGlnGluIleValAsnLeuSerLysGluAlaGlnLysPh | 857 |
| 985 | Qy | -----ArgAspSerProLeuGlnThrValGluArgGluLysThrLeuIle | 995 |
| 857 | Db | eAspSerSerLeuGlyAlaLeuLysThrGluLeuSerTyrLysThrGlnGluLeuGlnG | 877 |
| 996 | Qy | TCGCAGGCGAAGAAAGACACACCTAGGGAATATGACTCCCGAAGAAACACACTGAG | 1055 |
| 877 | Db | uLysThrArgGluValGlnGluArgLeuAsnGlnMetGluGlnLeuLysGluGlnLeuG | 897 |
| 1056 | Qy | AAATTACGTGGGAGAAAAGAAAGACCTAGGAGACGCATCGGAGAAAAAGAACCCCTGT | 1115 |
| 897 | Db | uAsn-----ArgAspSerProLeuGlnThrValGluArgGluLysThrLeuIle | 913 |
| 1116 | Qy | AAAGATGGATCGTCGAAGAGTAACATCTAATAAAGTTTGAAGAAAGAAATCTA | 1175 |
| 913 | Db | e-----ThrGluLysLeuGlnGlnThrLeuGluGluValLysThrLeuThrGlnGluLysAs | 932 |
| 1176 | Qy | AGATGTTGGCATGTCACAA-----AAGAATCATCTACAAAGCAAGATG | 1217 |
| 932 | Db | pAspLeuLysGlnLeuGlnGlnSerLeuGlnIleGluArgAspGlnLeuLysSerAspIle | 952 |
| 1218 | Qy | CCATGATCAGAGTCCTCATCAGATCCAAACAAACGGAAGAT----- | 1257 |
| 952 | Db | eHisAspThrValAsnMetAsnIleAspThrGlnGlnGlnLeuArgAsnAlaLeuGluSe | 972 |
| 1258 | Qy | -----GAAGAATAATCTTCAATCTCGGGTCTCTTGAAGTTCTGCAAGATTCAGTGGTA | 1313 |
| 972 | Db | rLeuLysGlnHisGlnGluThrIleAsnThrLeuLysSerLys----- | 986 |
| 1314 | Qy | TACCTGAGTCTATATATAAAAGTAATGACATAAATAAGAAGTAGAAGCTCCTCTAAAG | 1373 |
| 987 | Db | -----IleSerGluGluValSerArgAsnLe----- | 995 |
| 1374 | Qy | CCATCTGCCCTTCAAGCGCCATTGAAAGCAAAACTCTTTCCAA----- | 1416 |
| 995 | Db | uHisMetGluGluAsnThrGlyGluThrLysAspGluPheGlnGlnLysMetValGlyIle | 1015 |
| 1417 | Qy | -----ATAAGCCTTGAATGAAGATGAACAAACATTG----- | 1449 |
| 1015 | Db | eAspLysLysGlnAspLeuGluAlaLysAsnThrGlnThrLeuThrAlaAspValLysAs | 1035 |
| 1449 | Qy | ----- | 1449 |
| 1035 | Db | pAsnGluIleIleGluGlnGlnArgLysIlePheSerLeuIleGlnGluLysAsnGluIle | 1055 |
| 1450 | Qy | -----AAGCAGATCCGTGTTCC-----ACCGA | 1472 |
| 1055 | Db | uGlnGlnMetLeuGluSerValIleAlaGluLysGluGlnLeuLysThrAspLeuLysG | 1075 |
| 1473 | Qy | ATCCAAACAAAGACTATAGAATAATCTTGGATCTAGAGTCTCTCGAGACTGTTTA- | 1531 |
| 1075 | Db | uAsnIleGluMetThrIle-GluAsnGlnGluLeuArgLeuLeuGlyAspGluLeuL | 1095 |
| 1532 | Qy | -----CAGAAGGATTCGTGTTTACC-----GGCTACA | 1559 |
| 1095 | Db | ysLysGlnGlnGluIleValAlaGlnGluLysAsnHisAlaIleLysLysGluGlyGluL | 1115 |
| 1560 | Qy | CATCAAAAAGATAGATAAAAAATGGAAAAATAGAAGATCCCTATAAAGTGGTCT--- | 1616 |
| 1115 | Db | eUsrArgThrCysAspArgLeuAlaGlu-ValGluGluLysLeuLysGlnLysSerGln | 1134 |
| 1617 | Qy | -----TCTCAAGGCTAC | 1628 |
| 1135 | Db | GlnLeuGlnGluLysGlnGlnGlnLeuLeuAsnValGlnGluLwMetSerGluMetGln | 1154 |
| 1629 | Qy | CTCGGAATGAAATTTCTATTCACATAAAGCCCTAGAAATTCAGACATGCAAACTTTCA--- | 1685 |
| 1155 | Db | LysLysIleAsnGluIleGluAsnLeuLysAsnGluLeuLysAsnLysGluLeuThrLeu | 1174 |


```

Db      1790 leAspLysLeuArgGlylleValSerGluLysThrAspLysLeuSerAsnMetGlnLysA 1810
QY      3532 AACACATGTATCAAAACAGCAAGATAATGGAACACACCTGAACAGAGTCTCTAGT 3591
Db      1810 spLeuGluAsnSerAsnAlaLysLeuGlnGluLysIleGlnGlu-LeuLysAlaAsnGlu 1829
QY      3592 CAGAAATATTTCACATACAGCAAAATCTGGCTTCACAGCAATAGTAGTCACATA 3651
Db      1830 HisGlnLeuIleThrLeuLysLysAspValAsnGlnThrGlnLysValSerGluMet 1849
QY      3652 GAAAGCTGACACAAAGCAGATA----- 3675
Db      1850 GluGln---LeuLysLysGlnIleLysAspGlnSerLeuThrLeuSerLysLeuGluIle 1868
QY      3676 -----ACATGATATTCTTCGAGGAAATGCAACATCTCC 3717
Db      1869 GluAsnLeuAenLeuAlaGlnGluHisGluAsnLeuGluMetLysSerValMet 1888
QY      3718 TAAAGAGAAAAGAGGAGATATTAATTAACATACCATTTAAACCCGTATATCAATAT 3777
Db      1889 LysGluArgAsp-----AsnLeuArgVal---GluGluThrLeuLysLeu 1903
QY      3778 GAAAGAGAGAAGAGAAGAACTCATGAGACACAGCAGTAGAAATCTCTTTGGAGAA 3837
Db      1904 GluArgAspGlnLeuLysGluSerLeuGlnGluThrLysAlaArgAspLeuGluIleGln 1923
QY      3838 ACAACAGACAGACTTTACTCACACTCATGCTAGAGCCAGCTCTAGCATCACCTATGTTG 3897
Db      1924 GlnGluLeuLysThrAlaArgMetLeuSerLysGluHisLysGluThrValAspLysLeu 1943
QY      3988 -----AAATCTTACCAAGTCTGTGTCAGATACTATTTAGAGAAATTCATGT 3951
Db      1944 ArgGluLysIleSerGluLysThrIleGlnIleSerAspIleGln-LysAspLeuAspLys 1963
QY      3952 TTCTTCTCGAAGCTCAGCATTAATA-----AGTGTGAGATTAAGTGT 3999
Db      1963 sSerLysAspGluLeuGlnLysLysIleGlnGluLeuGlnLysLysLeuGlnLeu 1983
QY      4000 CCGAATGCAATAGCTGCACAGGATTCCTCATCCCTGATGTCAGCAGACTCAT---TC 4056
Db      1983 uArgValLys-----GluAspValAsnMetSerHisLysLys 1995
QY      4057 AATCAACACG-----AATCCGCTCTGC 4078
Db      1995 sIleAsnGluMetGluGlnLeuLysLysGlnPheGluProAsnTyrLeuLys 2012

RESULT 11
RA50_METJFA
ID_RA50_METJFA STANDARD; PRT: 1005 AA.
AC_Q58718;
DT_16-OCT-2001 (Rel. 40, Created)
DT_16-OCT-2001 (Rel. 40, Last sequence update)
DT_28-FEB-2003 (Rel. 41, Last annotation update)
DE_DNA double-strand break repair rad50 ATPase.
GN_RAD50 OR M1322.
OS_Methanococcus jannaschii.
OC_Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC_Methanocaldococcaceae; Methanocaldococcus.
OX_NCBI_TaxID=2190;
RN_[1]
RP_SEQUENCE FROM N.A.
RC_STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX_MEDLINE=96337999; PubMed=8688087;
RA_Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA_Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA_Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA_Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA_Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA_Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA_Cotton M.D., Roberts K.M., Hirst M.A., Kaine B.P., Borodovsky M.,
RA_Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus

```

```

RT_jannaschii";
RL_Science 273:1058-1073(1996).
CC_-1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC_rad50/mre11 complex possesses single-strand endonuclease activity
CC_and ATP-dependent double-strand-specific exonuclease activity.
CC_Rad50 provides an ATP-dependent control of mre11 by unwinding
CC_and/or repositioning DNA ends into the mre11 active site (By
CC_similarity).
CC_-1- SUBUNIT: Forms a complex with mre11 (By similarity).
CC_-1- SIMILARITY: Belongs to the SMC family. RAD50 subfamily.
CC-----
CC_This SWISS-PROT entry is copyright. It is produced through a collaboration
CC_between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC_the European Bioinformatics Institute. There are no restrictions on its
CC_use by non-profit institutions as long as its content is in no way
CC_modified and this statement is not removed. Usage by and for commercial
CC_entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC_or send an email to license@isb-sib.ch).
CC-----
DR_EMBL: U67572; AAB99331.1; -.
DR_PIR: A64465; A64465.
DR_TIGR: M1322; -.
DR_HAWAP: MF_00449; -.
DR_InterPro: IPR003593; AAA ATPase.
DR_InterPro: IPR003439; ABC transporter.
DR_InterPro: IPR007517; Rad50 zn_hook.
DR_InterPro: IPR003405; SMC_C.
DR_InterPro: IPR003395; SMC_N.
DR_Pfam: PF04423; Rad50 zn_hook; 1.
DR_Pfam: PF02483; SMC_C; 1.
DR_Pfam: PF02463; SMC_N; 1.
DR_Pfam: PF00006; ABC transporter; 1.
DR_SMART: SM00382; AAA; 1.
DR_NP_BIND: 32 39
FT_NP_BIND ATP (By SIMILARITY).
FT_DOMAIN 138 849 COILED COIL (POTENTIAL).
SQ_SEQUENCE 1005 AA; 119387 MW; 98BBB48173E789F3 CRC64;

Alignment Scores:
Pred. No.: 0.00478 Length: 1005
Score: 127.00 Matches: 192
Percent Similarity: 33.94% Conservative: 146
Best Local Similarity: 19.28% Mismatches: 291
Query Match: 1.79% Indels: 368
DB: 1 Gaps: 41

US-09-602-362E-22 (1-4115) x RA50_METJFA (1-1005)
QY 1402 CAAACTCTTCCAAATAGCCTTTGAATGAGATGACAAACATTGAAGCAGATCCGT 1461
Db 123 ArgAsnMetPheLeuAsnSerIleTyrIleLysGlnGlyLeuAlaLysPheLeuSer 142
QY 1462 GTTCCACCGAATCCAAACAAAGGACTATAGAAAATTTCTGGGATTCTAGAGTCTCTGG 1521
Db 143 LeuLysProSerGluLysLeuGluThrValAlaLysLeuLeuGlyIleAspGluPheGlu 162
QY 1522 AGACTGTTTACAGAGGATGTGTTTACAGGCTACATCAAAAGAAATAGATAAAA 1581
Db 163 Lys-CysTyrGlnLys-----MetGlyLeuIleValLysGluTyrGluLysAr 178
QY 1582 AATGAAAATAGACAGAGTCCCTAATAAAGTGCTTCTTGAAGGCTACCTCGGAATGAAT 1641
Db 178 gluGlu----- 180
QY 1642 TTCTATTCCATPAAAGCCTTAGAATGAGACATGCAAACTTTCAAGGAGCCTCCGGA 1701
Db 181 -----A-gileGluGlyLeuAsnTyrLysGluAsn----- 191
QY 1702 AGCCATCTGCTTCGAGCCTCCACATGAATCAAAAGTCTGTCCCAATAGAGCTTGAATG 1761
Db 192 -----TyrGluLysGluLeuLysAsnLysMetSer-----GlnLeuGluG 205
QY 1762 AAAAAATGAAAAC-----ATGGAGGCAGATGAG-----AACTCCCATCAGATCCAAACA 1812

```



```
QY 3621 TGTGGCTTCAGCAAGTAATGATCATGACATAGAAAGCTGACACAAAGAGAGATAAC--- 3677
Db 780 AspGluLeuAsnTyrAsnGluGluGluHisLysLysLysLysLysGluLeuTyrGluAsnLys 799
QY 3678 ---AATGATATCTATCTTCTGAGAGAAATGCAACATCTCTCTAAAGAGAGAAAGAGAG 3734
Db 800 ArgGlnGluLeuAspAsnValArgGluGlnLysThrGluLeu----- 813
QY 3735 AGATATTAATTAATACATACCATTTAAACCGTATATTCATATATGAAAGAGAGAGAA 3794
Db 814 GluThrGlyIleGluTyrLeuLysAspValGluSerLeuLysAlaArgLeuLysGlu 833
QY 3795 ACAGAACTCATGAGAACAGCAGTACAGAACTTCTTTGGAGAAACACAGACCAGACTTT 3854
Db 834 MetSerAsnLeuGluLysGluLysGluLysLeu----- 844
QY 3855 ACTCACACTATGCTAGAGCCAGCTGATGACATACCTATGTTGAAATCTTACCAAGTC 3914
Db 845 -----ThrllysPheValGluTyrLeuAspLysVal 854
QY 3915 TGTGTCACA-----GAATACCTATTTTGAAGAAATTCATG----- 3950
Db 855 ArgArgIlePheGlyArgAsnGlyPheGlnAlaTyrLeuArgGluLysTyrValProLeu 874
QY 3951 -----TTTCTCTCTAGCCCTACAGC 3971
Db 875 IleGlnLysTyrLeuAsnGluAlaPheSer 884

RESULT 12
MYSJ_DICDI
ID MYSJ_DICDI STANDARD; PRT; 2245 AA.
AC PS4697;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Mysin IO heavy chain.
GN MYOJ.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=96215148; PubMed=8636147;
RA Hammer J.A. III, Jung G.;
RT "The sequence of the dictyostelium myo J heavy chain gene predicts a
RT novel, dimeric, unconventional myosin with a heavy chain molecular
RT mass of 258 kDa."
RL J. Biol. Chem. 271:7120-7127(1996).
RN [2]
RP SEQUENCE OF 1-1021 FROM N.A.
RX MEDLINE=97039016; PubMed=8884597;
RA Peterson M.D., Urioste A.S., Titus M.A.;
RT "Dictyostelium discoideum myoJ: a member of a broadly defined myosin
RT V class or a class XI unconventional myosin?";
RL J. Muscle Res. Cell Motil. 17:411-424(1996).
RN [3]
RP SEQUENCE OF 182-298 FROM N.A.
RX MEDLINE=95023928; PubMed=7937787;
RA Titus M.A., Kuspa A., Loomis W.F.;
RT "Discovery of myosin genes by physical mapping in Dictyostelium.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9446-9450(1994).
CC -!- FUNCTION: Mysin is a protein that binds to actin and has ATPase
CC activity that is activated by actin.
CC -!- SUBUNIT: Homodimer.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 dilute domain.
CC -!- SIMILARITY: Contains 3 IQ domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC -----
DR EMBL; U42409; AAA85186.1; -.
DR PIR; T18278; T18278.
DR HSP; P08799; IAMD.
DR DictyBase; DDB0185050; myoJ.
DR InterPro; IPR002710; DIL.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin head.
DR InterPro; IPR004009; Myosin_N.
DR Pfam; PF01843; DIL; 1.
DR Pfam; PF00612; IQ; 6.
DR Pfam; PF00063; myosin head; 2.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD003376; DIL; 1.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 3.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS0096; IQ; 3.
KW Myosin; ATP-binding; Actin-binding; Calmodulin-binding; Repeat;
KM Coiled coil.
FT DOMAIN 1 809 MYOSIN HEAD-LIKE.
FT DOMAIN 824 851 IQ 1.
FT DOMAIN 872 901 IQ 2.
FT DOMAIN 943 972 IQ 3.
FT DOMAIN 973 1812 COILED COIL (POTENTIAL).
FT DOMAIN 2060 2220 DILUTE.
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 669 749 ACTIN-BINDING.
FT CONFLICT 191 191 L -> F (IN REF. 2).
FT CONFLICT 284 284 A -> T (IN REF. 2).
FT CONFLICT 291 291 G -> R (IN REF. 2).
FT CONFLICT 332 347 NKSGCFEIGVSDEEH -> IEWPFELKYVRMS (IN REF. 2).
FT CONFLICT 550 550 N -> K (IN REF. 2).
FT CONFLICT 865 866 HH -> QQ (IN REF. 2).
SQ SEQUENCE 2245 AA; 258478 MW; 61585EFID1AB45BE CRC64;

Alignment Scores:
Pred. No.: 0.00737 Length: 2245
Score: 125.50 Matches: 225
Percent Similarity: 37.59% Conservative: 181
Best Local Similarity: 20.83% Mismatches: 339
Query Match: 1.77% Indels: 335
DB: 1 Gaps: 50

US-09-602-362E-22 (1-4115) x MYSJ_DICDI (1-2245)
QY 886 GAAAGACATCTGGAAGTTGCAAGCTGACGAGAAAACACCTAGGAAATTCAGATCCTG 945
Db 969 GluAlaArgSerLeuArgThrValGlnGluLysAsnLys-----Leu 983
QY 946 CAAAAAATCTGAGAAATTTAC-----GGCCAGCAAGAGAGAC 987
Db 984 GlnGluLysLeuGluGluLeuGlnTyrArgLeuThrSerGluAlaLysArgLysGlnGln 1003
QY 988 CTAGAGATCGCAGGAGAGAAAGACACACCTAGGGAATATGAGTCCCGAAAAGAA 1047
Db 1004 LeuGluAspGlnLysValLysSerAspThrThrThrSerGluLeuSerSerAsnAsnAsp 1023
QY 1048 ACACCTGAGAAATTCAGTGGGAGAGAAAGAGACACCTAGGAGACGATGGGAGAGAAA 1107
Db 1024 -HisLeuGluLeuGlnLeuSerGluLeuGln----- 1033
QY 1108 AACCTGTAAAGATCGGTCGCAAGAGTAGTACATCTAATAACTAAAGTTTGAAGAA 1167
Db 1034 -----LeuLysTyrGlnGluLeuAspAsnSerAsnGln-SerSerGlnL 1048
```

| | | | |
|----|------|---|------|
| QY | 1168 | GAATCTTAAGATGTTGCATGTCACAAAAAGAAATCATCTACAAAGCAAGTGCCATGATCAG | 1221 |
| Db | 1048 | euGlnLeuSerGluCysLeuSerLysLeuGluGlnThrGlnLeuAaspHisSer- | 1067 |
| QY | 1228 | AGGTCCCATCAGATCCAAACAAGGAAGATGAAGAATATTCTTTGATCTTCGGGTCTCTT | 1287 |
| Db | 1068 | -----SerLysLeuAsnLysLysLeuGluLysAspLeuSerAspGlnHisAspSerI | 1085 |
| QY | 1288 | TGAAGTTCTCCAAAGATCAAGTGGTATACCTGAGTCTTATATATAAAAGTAATAGATAAA | 1347 |
| Db | 1085 | leGlu-----LysLeuGln-----SerGlnPheA | 1093 |
| QY | 1348 | ATAAGAGTAGAAAGCCCTCTAAAGCCATCTGCCTTCRAAGCCGCCATTGAAAGCAAAAC | 1407 |
| Db | 1093 | sn-GluThrGlnGlnLeuGlnGlnPheLysGlnGlnSerGluLeuSerSerLy | 1112 |
| QY | 1408 | TCCTTCCAAATAAGCCCTTCAATGAAGAAATCAACAAACATTGAAGCAGATCCGTGTCTCC | 1467 |
| Db | 1112 | sLeuSerLysThrThrGlnGlnLeuAspPheAsnLysGln-GluPheAspArgLeuSerG | 1132 |
| QY | 1468 | ACCGAATCCAAACA----- | 1481 |
| Db | 1132 | lnGluArgAspThrAspAsnThrAsnGlnLeuGluIleGlnGlnLeuLysLysAlaA | 1152 |
| QY | 1482 | -----AAGCATCTATAAGAAAAATCTTGGGATTTCTAGAGTCTCTCGAGACTG | 1527 |
| Db | 1152 | snSerThrLeuGluGluAspTyrPheSerLeuSerGlyIleArgAspAsnLeuGlu- | 1170 |
| QY | 1528 | TTTACAGAAGGATGTGTTTACCAGGCTACACATCAAAAGAAATAGATAAAAAAATGGA | 1587 |
| Db | 1171 | -----ArgGlnValLeu-----GluLeuA | 1177 |
| QY | 1588 | AAATAGAGAGTCCCTAATAAAGTGTCTTCTGAAGGCTACCTCGGAATGAAATTTCTAT | 1647 |
| Db | 1177 | rgAspGluAsnGlnLeuIleLysGluArgLeuAsp----- | 1188 |
| QY | 1648 | TCCACTAAAGCCTAGAAATTGAAGCATCAAAACTTTCAAAGCGAGCCTCCGGGAAGCCAT | 1707 |
| Db | 1189 | -----SerLeuGlyGlnGlnSer-SerGlnPheGlnSer-----GlyAlaAla | 1202 |
| QY | 1708 | CTGCTTCGAGCCTCCACTGAA-----ATCAAAAGTCTGTCCCAATAAGCCTTGAATG | 1761 |
| Db | 1203 | LeuGluLysGlnGlnLeuGluGlnLeuValGlnGlnGlnSerGluGlnLeuIleLysLeu | 1222 |
| QY | 1762 | AAAAATCAAAAACATGAGGAGCAGATGAGAATCCCATCAGAATCCAAACAAAGACTATA | 1821 |
| Db | 1223 | SerSerGluLysLeuGlySerGluGluGluAlaLysLysGlnIleAsnGlnLeuGluLeu | 1242 |
| QY | 1822 | AGAAAATTCGGGATACTGGAGTCTCTGTAGACTGTTTTCAGNAGGATTGTCTTTACCA | 1881 |
| Db | 1243 | Glu-----LeuThr | 1245 |
| QY | 1882 | AGCTCGCTCAAAA-----AGATAGATAAAATAATGCAAAATTAGAAGGTCCTCGTTAA | 1938 |
| Db | 1246 | AspHisLysSerLysLeuGlnIleGln----- | 1254 |
| QY | 1939 | AGTGGTCTCTGAGGCTAACT-----GGGAATGAAAGTCTTATTCACATAAGCCCTAG | 1992 |
| Db | 1255 | -----LeuGlnLeuThrGluGlnSerAsnGluLys-IleLysLysLeuLysGlyLy | 1271 |
| QY | 1993 | AATTGATCAGATGCAAACTTCAAAGCAGGCTCCCGAGAGCCATCTGCCTCGAGCCTCC | 2052 |
| Db | 1271 | sLeuGluGluTyrGlnAspGluLysLysGlnLeuGlnGln-----GluLeuGluArgI | 1289 |
| QY | 2053 | ATTGAAATCAAAGTCTGTCCAAATAAACCTTGGAAATTAGAAATGAACAACATTGAGGC | 2112 |
| Db | 1289 | eLysGlnSerLys-----GlnSerValGluAs | 1298 |
| QY | 2113 | AGATGAGACTCC-----CATCAATCCAAACAAGGACTATGCAAGAAAGTTCTTGGAT | 2166 |
| Db | 1298 | pGluLysAsnSerLeuIleThrGlnLeuThrThrValLysPheGluSerThr----- | 1315 |
| QY | 2167 | TCTGGAGTCTCTGAGACTGTTTCCAGAAGGATGGTGTTTACCAGGCTACACATCAAAA | 2226 |

| | | | | |
|----|------|---|---|------|
| Db | 1316 | | -----GlnValSerThrAsnValSerHisGlnLys | 1322 |
| Qy | 2227 | AGAAATAGATAAAATAATGGAAA | -----ATPAGAAGAGCTCTCTGATGAATAGGTTTTCTGA | 2280 |
| Db | 1325 | sGluLysIleThrThrLeuLysSerThrIleGluGlu | -----LeuAsnLysSerIle | 1342 |
| Qy | 2281 | AGCTCCCTCAGAAATGAAGCTTCTATCCACTAAAGCCITGAATTGATGCATGCACAACT | 2340 | |
| Db | 1342 | eGlyLysLeuGln | ----- | 1346 |
| Qy | 2341 | TTCAAAAGCAGCGCTCCCGAGAGCCATCTGCTTCGAGCTGCATTTGAATAAGAAAGTCTGT | 2400 | |
| Db | 1347 | -AlaGluGlnLysAsnLysAspGlu | ----- | 1355 |
| Qy | 2401 | TCCAAATAAACCTTGGNAITTAAGAAATGAACAAACATTCAGACACATCAGATTTCCCTTCAG | 2460 | |
| Db | 1356 | -----IleArgLysIleGlnPheGluLeuAs | 1364 | |
| Qy | 2461 | AATCAAAACAAAGAAAGGTTTGAGAAAATTCGGGATTCCTGAAGTCTCCGTGGAGCTGTTCA | 2520 | |
| Db | 1364 | nasPginLysGln | ----- | 1368 |
| Qy | 2521 | CAGAAGATCGTGTGTACCCAGCGTACACACAAAAGAAAGGATAAATAAGTGGAAAATT | 2580 | |
| Db | 1369 | -----GlnPheThrArgGlnThrLys | -----GluPh | 1377 |
| Qy | 2581 | AGAAGATTCTAGCCTATCAAAATCTTGGT | -----ACAGTTTCAT | 2620 |
| Db | 1377 | eSerAspLeuGlnSerGlnGlnSerIleAspArgProLysSerGluIleThrIleHis | 1397 | |
| Qy | 2621 | -CTTGTGAAAGCGAAGGAACTTCAAAGACACTGTGAACACGTACAGAAAT | 2676 | |
| Db | 1397 | rLeuLysArgThrAsnGlnThrLeuLysSerAspPheGluArgValGlnGlnSerLeuLys | 1417 | |
| Qy | 2677 | AACATCAAAAGAGTTTTCTGTACTGAAAAGAACTGTCCAGA | 2730 | |
| Db | 1417 | sGlnGlnLysArgAspCysGlnGlnThrLysAspThrIleAsnArgLeuGluAsnGluVal | 1437 | |
| Qy | 2731 | AAATCAGATAGAGAACCAAAAGTTAAATCGGAACAGAGCTGTGCAGTGTAGATTGCATT | 2790 | |
| Db | 1437 | lLysGlnLeuThrGlnLeuLysGluArgPheGluAsnGluPhePheValAlaLysGluGln | 1457 | |
| Qy | 2791 | AAACCAAGAAAGAGACAGAGAAATGCCCATATATTAAAGAAAAAATTAGGAGAAATTA | 2850 | |
| Db | 1457 | nasSerAsnGlnThrGlnGlnSerVal | -----TyrLeuLysGluValThrThrGlnMetGln | 1476 |
| Qy | 2851 | GAAGAAATCGAGACAGACATAGAAGAGTTAGAAGTGAACACAA | 2904 | |
| Db | 1476 | nGlnAsn-GlnSerArgIleGluArgGluLeuGluLysGlnHisIleThrArgI | 1496 | |
| Qy | 2905 | TCTCAGATACAGATAAGATTGAAGGTGTAGAAAGTAATTTGAATAGGTTTCTCACTC | 2964 | |
| Db | 1496 | leAspAspGluArgAspGluLeuLys | -----LysGlnLeuThrGlnLeuGlnGlnGlnH | 1514 |
| Qy | 2965 | ATGAAATG | -----AAAATTATCTTACATCAAAAT | 3006 |
| Db | 1514 | isGluGlnSerSerThrGlnLeuLeuLeuAlaGlnAsnGluLeuGluArgLeuLysL | 1534 | |
| Qy | 3007 | AGGAATTGCCATGCAAA | ----- | 3024 |
| Db | 1534 | ysGluLeuLysTyrLysGluArgGlyHisGluThrSerLysGlnGlnAspGlnPheAsnM | 1554 | |
| Qy | 3025 | CTGGAATAGCCACATAAACACCAATCCAGGAAAAAGGAAAAATAAATCTTTGACGACTTAA | 3084 | |
| Db | 1554 | etGluIleGlnSerLeuArgIleThr-AsnAsnAspGlnLeuLysSerLeuGln | 1571 | |
| Qy | 3085 | GATTTTAAAGAAAGATGCTCAACTTAGATGAC | 3120 | |
| Db | 1572 | AspTyrGluGlnGluLysLysLysLeuLysAspLysLeuSerSerSerLysGlnGluAla | 1591 | |
| Qy | 3121 | AAAACCTGAAAGGAATCATTAATAAAAGGATCTCAATATGTGGCAGCTAAAGTTCTGA | 3180 | |

| | | | | | | | |
|----|----------|------|---|----|------|---|------|
| FT | CONFLICT | 695 | R -> S (IN REF. 1 AND 3). | QY | 2449 | ATTTCCTTTCAGAAATCAAAACAAAGAGGTTGAGAAATTTCTGGATTTCTGAAGTCTCCG | 2508 |
| FT | CONFLICT | 736 | ENSTTTT -> RKNHHD (IN REF. 3). | Db | 767 | LysValTyrLysIleGlyAsnThrLysLeuPhePheLysAlaGlyValLeuAlaAsp--- | 785 |
| FT | CONFLICT | 756 | E -> R (IN REF. 1 AND 3). | QY | 2509 | TGGACTGTTTTCACAGAAGAT--- | 2535 |
| FT | CONFLICT | 773 | NTLFFKAGVLA -> ILTVFKLEYS (IN REF. 1). | Db | 786 | ---LeuGluLysGlnLysAspValLysLeuAsnIleMetIleLysLeuThrAlaThr | 804 |
| FT | CONFLICT | 793 | KL -> NV (IN REF. 1). | QY | 2536 | TACCACGGCTACACA--- | 2559 |
| FT | CONFLICT | 794 | N -> T (IN REF. 1). | QY | 805 | IleArgGlyTyrThrValArgLysGluIleThrTyrHisLeuGlnLysLeuLysLysThr | 824 |
| FT | CONFLICT | 896 | N -> NSQITKINTNITETPQSYTIGERPMPVICGN (IN REF. 1). | Db | 2560 | AGATAAAATAGTGGAAATAGAAATTCACCTACCTATCAAAATCT--- | 2613 |
| FT | CONFLICT | 900 | N -> I (IN REF. 1). | QY | 825 | ArgValIleGlyAsnThrPheArgLeuTyrAsnArgLeuValLysGluAspProTyrPhe | 844 |
| FT | CONFLICT | 906 | N -> K (IN REF. 1). | QY | 2614 | AGTTTCATCTTGTGAAAGGCAAGGAACTT--- | 2664 |
| FT | CONFLICT | 911 | N -> K (IN REF. 1). | Db | 845 | Asn--LeuPheIleArgIleLysProLeuLeuThrSerSerAsnAspMetThrArgThrL | 864 |
| FT | CONFLICT | 915 | NESLNVRKTSSETLQ -> RIAIKILPAINIT (IN REF. 1). | QY | 2665 | ACAGGAAATGGAACAAATGAAAGAGTTTGTGTACTGAAAGAAAGAACTCTCAGAACAA | 2724 |
| FT | CONFLICT | 930 | DDUSE -> MTLFL (IN REF. 1). | Db | 864 | ysLysPheAsnGluGlnIleAsnLysLeuLysAsnAspLeuGlnGluMetGluSerLysL | 884 |
| FT | CONFLICT | 934 | EAH -> RKI (IN REF. 1). | QY | 2725 | AGAAATAATACACAGTAGAGAACCAAGAGTTAAATGGGAACAGAGTCTGACGTGTAGAT | 2784 |
| FT | CONFLICT | 951 | S -> C (IN REF. 1). | Db | 884 | ysLysPheLeuGluLysAsnGlnLys--- | 893 |
| FT | CONFLICT | 955 | L -> D (IN REF. 1). | QY | 2785 | TGACTTAAACCAAGAAAGAGAGAGAAATGCCGATATATTAAAGAAATAAGGAA | 2844 |
| FT | CONFLICT | 958 | L -> D (IN REF. 1). | Db | 894 | ---ThrValAsnGluLeuGluAsnThrGlnAspLeuLeuAsnGlnGluAla | 911 |
| FT | CONFLICT | 1002 | C -> S (IN REF. 1). | QY | 2845 | GAATTAGAGATCGAGAGCAGCATAGAAAGATTAGAGTGAACACAACTTGAAGGC | 2904 |
| FT | CONFLICT | 1049 | C -> S (IN REF. 1). | Db | 911 | snLeuArgLysAsnGluSerLeuLeuAsnArgValLysThrSerSerGluThrLeuGln- | 930 |
| FT | CONFLICT | 1049 | C -> S (IN REF. 1). | QY | 2905 | TCTCAGATACAGATAGAAATGGAAGTGTAGAAAGTAATTTGAATAGGTTCT- | 2958 |
| FT | CONFLICT | 1056 | M -> I (IN REF. 1). | Db | 931 | LysGlnPheAspAspLeu- | 940 |
| FT | CONFLICT | 1060 | A -> E (IN REF. 1). | QY | 2959 | ---CAACTCATGAAATGAAATTTCTTACATGAAATTCATGCTTGAAGAAAGAAATTC | 3015 |
| FT | CONFLICT | 1085 | V -> C (IN REF. 1). | Db | 941 | AspGluIleSerArgGluLysLeu- | 957 |
| FT | CONFLICT | 1123 | L -> S (IN REF. 1). | QY | 3016 | CATGCAAACTGGAATAGCCACACATAACCAATCCAGGAAAGGAAATAATTTTG | 3075 |
| FT | CONFLICT | 1133 | KSN -> NLI (IN REF. 1). | Db | 958 | HisGlnLysIleGln- | 966 |
| FT | CONFLICT | 1146 | RETKEQK -> TRKKEQDK (IN REF. 1). | QY | 3076 | AGGACTTAAGATTTTAAAGAAAGATGCTGAACCTAGATGACCTTAAACTGAAAGGAA | 3135 |
| FT | CONFLICT | 1159 | SKI -> ELKV (IN REF. 1). | Db | 967 | ---ThrIleArgGluLysGluAlaThrLeuGluLysLeuHisSerLysAsnAsn | 983 |
| FT | CONFLICT | 1181 | LE -> WK (IN REF. 1). | QY | 3136 | TCATTACTTAAAGGATCTCAATATGTGGGAGCTTAAAGTTCTGATAGTGAAGAAAT | 3195 |
| FT | CONFLICT | 1185 | LSQETSLNQYLNRISG -> CHRKLKSLKQNR (IN REF. 1). | Db | 984 | GluLeuIleLysGlnIleSer- | 1000 |
| FT | CONFLICT | 1204 | P -> S (IN REF. 1). | QY | 3196 | GCTCATTCTAAATTCAGGAAAGAACAGACAAAGAACTAGAGCAAGAAATTTGAATACAC | 3255 |
| FT | CONFLICT | 1224 | D -> H (IN REF. 1). | Db | 1001 | SerSerGlnSerLeuIleLysGluSerLysLeuLysGluAsnGluLysLys- | 1018 |
| FT | CONFLICT | 1228 | E -> Q (IN REF. 1). | QY | 3256 | ATCCTGACTGCTTGGTGTACAGCCATGATCAATTTGTGACATCAAGAAAGAAAGTAAAGAA | 3315 |
| FT | CONFLICT | 1228 | E -> Q (IN REF. 1). | Db | 1019 | -----ArgLeuLysAspValIleAsnSer--LysGluGluGluGluLysLys | 1033 |
| FT | CONFLICT | 1253 | POKESDKMLLE -> LTKSLILTNGNAS (IN REF. 1). | QY | 3316 | CCTGCTCCCATTTGAGGAGATGCTGTTTCAAGAAAGTGAATGTTGTGAGAGTAC | 3375 |
| FT | CONFLICT | 1311 | D -> V (IN REF. 1). | Db | 1033 | er-----PheAsnAspLysSer- | 1039 |
| FT | CONFLICT | 1323 | DLLKOLDHYTKVEMLN -> SEARSILYKSGNVD (IN REF. 1). | QY | 3376 | GATATTAAACATGAGTGTCCATCCACCTTCTTGAGCTCAAGGAAATCCAAACCTAA | 3435 |
| FT | CONFLICT | 1400 | MISSING (IN REF. 1). | Db | 1040 | -----SerSerGluGluAspLeuAspIleL | 1048 |
| FT | CONFLICT | 1400 | TIQLQNEQSRNG -> NTTANGTKFKEM (IN REF. 1). | QY | 3436 | AAATTATCTCAATTAGCAGGAGATGCTTAAGAGAAATACATTGCTTTCAGAACATCACAA | 3495 |
| FT | CONFLICT | 1454 | FODE -> LMM (IN REF. 1). | | | | |
| FT | CONFLICT | 1757 | D -> E (IN REF. 1). | | | | |
| FT | CONFLICT | 1777 | R -> T (IN REF. 1). | | | | |
| FT | CONFLICT | 1788 | S -> D (IN REF. 1). | | | | |
| FT | CONFLICT | 1825 | S -> W (IN REF. 1). | | | | |
| FT | CONFLICT | 1882 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1902 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1904 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1928 | AA; 223634 MW; 6F54C7611F43DC9F CRC64; | | | | |
| FT | CONFLICT | 1704 | MISSING (IN REF. 1). | | | | |
| FT | CONFLICT | 1737 | TIQLQNEQSRNG -> NTTANGTKFKEM (IN REF. 1). | | | | |
| FT | CONFLICT | 1754 | FODE -> LMM (IN REF. 1). | | | | |
| FT | CONFLICT | 1777 | D -> E (IN REF. 1). | | | | |
| FT | CONFLICT | 1788 | R -> T (IN REF. 1). | | | | |
| FT | CONFLICT | 1825 | S -> D (IN REF. 1). | | | | |
| FT | CONFLICT | 1882 | S -> W (IN REF. 1). | | | | |
| FT | CONFLICT | 1902 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1904 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1928 | AA; 223634 MW; 6F54C7611F43DC9F CRC64; | | | | |
| FT | CONFLICT | 1698 | MISSING (IN REF. 1). | | | | |
| FT | CONFLICT | 1725 | TIQLQNEQSRNG -> NTTANGTKFKEM (IN REF. 1). | | | | |
| FT | CONFLICT | 1754 | FODE -> LMM (IN REF. 1). | | | | |
| FT | CONFLICT | 1777 | D -> E (IN REF. 1). | | | | |
| FT | CONFLICT | 1788 | R -> T (IN REF. 1). | | | | |
| FT | CONFLICT | 1825 | S -> D (IN REF. 1). | | | | |
| FT | CONFLICT | 1882 | S -> W (IN REF. 1). | | | | |
| FT | CONFLICT | 1902 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1904 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1928 | AA; 223634 MW; 6F54C7611F43DC9F CRC64; | | | | |
| FT | CONFLICT | 1698 | MISSING (IN REF. 1). | | | | |
| FT | CONFLICT | 1725 | TIQLQNEQSRNG -> NTTANGTKFKEM (IN REF. 1). | | | | |
| FT | CONFLICT | 1754 | FODE -> LMM (IN REF. 1). | | | | |
| FT | CONFLICT | 1777 | D -> E (IN REF. 1). | | | | |
| FT | CONFLICT | 1788 | R -> T (IN REF. 1). | | | | |
| FT | CONFLICT | 1825 | S -> D (IN REF. 1). | | | | |
| FT | CONFLICT | 1882 | S -> W (IN REF. 1). | | | | |
| FT | CONFLICT | 1902 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1904 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1928 | AA; 223634 MW; 6F54C7611F43DC9F CRC64; | | | | |
| FT | CONFLICT | 1698 | MISSING (IN REF. 1). | | | | |
| FT | CONFLICT | 1725 | TIQLQNEQSRNG -> NTTANGTKFKEM (IN REF. 1). | | | | |
| FT | CONFLICT | 1754 | FODE -> LMM (IN REF. 1). | | | | |
| FT | CONFLICT | 1777 | D -> E (IN REF. 1). | | | | |
| FT | CONFLICT | 1788 | R -> T (IN REF. 1). | | | | |
| FT | CONFLICT | 1825 | S -> D (IN REF. 1). | | | | |
| FT | CONFLICT | 1882 | S -> W (IN REF. 1). | | | | |
| FT | CONFLICT | 1902 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1904 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1928 | AA; 223634 MW; 6F54C7611F43DC9F CRC64; | | | | |
| FT | CONFLICT | 1698 | MISSING (IN REF. 1). | | | | |
| FT | CONFLICT | 1725 | TIQLQNEQSRNG -> NTTANGTKFKEM (IN REF. 1). | | | | |
| FT | CONFLICT | 1754 | FODE -> LMM (IN REF. 1). | | | | |
| FT | CONFLICT | 1777 | D -> E (IN REF. 1). | | | | |
| FT | CONFLICT | 1788 | R -> T (IN REF. 1). | | | | |
| FT | CONFLICT | 1825 | S -> D (IN REF. 1). | | | | |
| FT | CONFLICT | 1882 | S -> W (IN REF. 1). | | | | |
| FT | CONFLICT | 1902 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1904 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1928 | AA; 223634 MW; 6F54C7611F43DC9F CRC64; | | | | |
| FT | CONFLICT | 1698 | MISSING (IN REF. 1). | | | | |
| FT | CONFLICT | 1725 | TIQLQNEQSRNG -> NTTANGTKFKEM (IN REF. 1). | | | | |
| FT | CONFLICT | 1754 | FODE -> LMM (IN REF. 1). | | | | |
| FT | CONFLICT | 1777 | D -> E (IN REF. 1). | | | | |
| FT | CONFLICT | 1788 | R -> T (IN REF. 1). | | | | |
| FT | CONFLICT | 1825 | S -> D (IN REF. 1). | | | | |
| FT | CONFLICT | 1882 | S -> W (IN REF. 1). | | | | |
| FT | CONFLICT | 1902 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1904 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1928 | AA; 223634 MW; 6F54C7611F43DC9F CRC64; | | | | |
| FT | CONFLICT | 1698 | MISSING (IN REF. 1). | | | | |
| FT | CONFLICT | 1725 | TIQLQNEQSRNG -> NTTANGTKFKEM (IN REF. 1). | | | | |
| FT | CONFLICT | 1754 | FODE -> LMM (IN REF. 1). | | | | |
| FT | CONFLICT | 1777 | D -> E (IN REF. 1). | | | | |
| FT | CONFLICT | 1788 | R -> T (IN REF. 1). | | | | |
| FT | CONFLICT | 1825 | S -> D (IN REF. 1). | | | | |
| FT | CONFLICT | 1882 | S -> W (IN REF. 1). | | | | |
| FT | CONFLICT | 1902 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1904 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1928 | AA; 223634 MW; 6F54C7611F43DC9F CRC64; | | | | |
| FT | CONFLICT | 1698 | MISSING (IN REF. 1). | | | | |
| FT | CONFLICT | 1725 | TIQLQNEQSRNG -> NTTANGTKFKEM (IN REF. 1). | | | | |
| FT | CONFLICT | 1754 | FODE -> LMM (IN REF. 1). | | | | |
| FT | CONFLICT | 1777 | D -> E (IN REF. 1). | | | | |
| FT | CONFLICT | 1788 | R -> T (IN REF. 1). | | | | |
| FT | CONFLICT | 1825 | S -> D (IN REF. 1). | | | | |
| FT | CONFLICT | 1882 | S -> W (IN REF. 1). | | | | |
| FT | CONFLICT | 1902 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1904 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1928 | AA; 223634 MW; 6F54C7611F43DC9F CRC64; | | | | |
| FT | CONFLICT | 1698 | MISSING (IN REF. 1). | | | | |
| FT | CONFLICT | 1725 | TIQLQNEQSRNG -> NTTANGTKFKEM (IN REF. 1). | | | | |
| FT | CONFLICT | 1754 | FODE -> LMM (IN REF. 1). | | | | |
| FT | CONFLICT | 1777 | D -> E (IN REF. 1). | | | | |
| FT | CONFLICT | 1788 | R -> T (IN REF. 1). | | | | |
| FT | CONFLICT | 1825 | S -> D (IN REF. 1). | | | | |
| FT | CONFLICT | 1882 | S -> W (IN REF. 1). | | | | |
| FT | CONFLICT | 1902 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1904 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1928 | AA; 223634 MW; 6F54C7611F43DC9F CRC64; | | | | |
| FT | CONFLICT | 1698 | MISSING (IN REF. 1). | | | | |
| FT | CONFLICT | 1725 | TIQLQNEQSRNG -> NTTANGTKFKEM (IN REF. 1). | | | | |
| FT | CONFLICT | 1754 | FODE -> LMM (IN REF. 1). | | | | |
| FT | CONFLICT | 1777 | D -> E (IN REF. 1). | | | | |
| FT | CONFLICT | 1788 | R -> T (IN REF. 1). | | | | |
| FT | CONFLICT | 1825 | S -> D (IN REF. 1). | | | | |
| FT | CONFLICT | 1882 | S -> W (IN REF. 1). | | | | |
| FT | CONFLICT | 1902 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1904 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1928 | AA; 223634 MW; 6F54C7611F43DC9F CRC64; | | | | |
| FT | CONFLICT | 1698 | MISSING (IN REF. 1). | | | | |
| FT | CONFLICT | 1725 | TIQLQNEQSRNG -> NTTANGTKFKEM (IN REF. 1). | | | | |
| FT | CONFLICT | 1754 | FODE -> LMM (IN REF. 1). | | | | |
| FT | CONFLICT | 1777 | D -> E (IN REF. 1). | | | | |
| FT | CONFLICT | 1788 | R -> T (IN REF. 1). | | | | |
| FT | CONFLICT | 1825 | S -> D (IN REF. 1). | | | | |
| FT | CONFLICT | 1882 | S -> W (IN REF. 1). | | | | |
| FT | CONFLICT | 1902 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1904 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1928 | AA; 223634 MW; 6F54C7611F43DC9F CRC64; | | | | |
| FT | CONFLICT | 1698 | MISSING (IN REF. 1). | | | | |
| FT | CONFLICT | 1725 | TIQLQNEQSRNG -> NTTANGTKFKEM (IN REF. 1). | | | | |
| FT | CONFLICT | 1754 | FODE -> LMM (IN REF. 1). | | | | |
| FT | CONFLICT | 1777 | D -> E (IN REF. 1). | | | | |
| FT | CONFLICT | 1788 | R -> T (IN REF. 1). | | | | |
| FT | CONFLICT | 1825 | S -> D (IN REF. 1). | | | | |
| FT | CONFLICT | 1882 | S -> W (IN REF. 1). | | | | |
| FT | CONFLICT | 1902 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1904 | FWK -> NSGKELDADDL (IN REF. 1). | | | | |
| FT | CONFLICT | 1928 | AA; 223634 MW; 6F54C7611F43DC9F CRC64; | | | | |
| FT | CONFLICT | 1698 | MISSING (IN REF. 1). | | | | |
| FT | CONFLICT | 1725 | TIQLQNEQSRNG -> NTTANGTKFKEM (IN REF. 1). | | | | |
| FT | CONFLICT | 1754 | FODE | | | | |

```
Db 1048 ysLeuValThrLeuGluysnCyS-----AsnIlealaM 1060
QY 3496 AGAaACAGCTGAaAaCAGTCTCAATGAGaAGCAGaACATATATCAaAGCAAGa 3555
Db 1060 etSerArgLeuGlnSerLeuValThrGluAsnSerAspLeuArgSerLysAsnGluAsnP 1080
QY 3556 TAAaTGAAaCAa-----ACaCTGaACAGCAGaGCTCTAGCTCAaGAATATTT 3603
Db 1080 heLysLysGluLysAlaLaLeuAsnAsnGlnLysAsnLysGluSerGluLeuLeuL 1100
QY 3604 CAATaCAAGCAaAAaATGTGGCTTCAAGCAaATGATGATGACATAGaAGCTGCAa 3663
Db 1100 ysMetLysGluLys-----IleAspAsnHisLysLysGluLeuAlaThrP 1115
QY 3664 CAaAGCAGATaCAaATGATATCTTCTTGAGAGGAATGCAACATCTCTCAaAG 3723
Db 1115 heSerLysGlnArgAspAlaValSerGluHisGlyLysIleThrAlaGluLeuLysG 1135
QY 3724 AGaAAaAa-----AGGAGATATTAATTAATTAATTAATTAATTAATTAATTA 3771
Db 1135 luThrArgIleGlnLeuThrGluTyLysSerAsnTyLysIleLysGluLutyrS 1155
QY 3772 CAATaTGAaAAGaAGaAGaAaCAaAaCTCATGAGaAaCAaAGCAGTAACTCTTT 3831
Db 1155 er-----AsnPhGlnArgGluThrLysGluGlnGlnLysLys-ArgAsnSerLeu 1172
QY 3832 GGAGaAaCAaAGaCAGaCTTTACTCAGaCTCATGCTAGAGCCAGCTAGAGATCA--- 3888
Db 1173 ValGluSerLeuAsnAspSerLysIleLys---GluLeuGluAlaArgLeuSerGlnGlu 1191
QY 3889 ---CCTATGTTGAaAaCTTCAaAaAGTGTGTGTCAGaAATATCTATT 3933
Db 1192 IleSerLeuAsnGlnTyLysLeuAsnLysArgIleSerGlyAsnSerVal 1207

RESULT 14
GOBL_HUMAN
ID GOBL_HUMAN STANDARD; PRT; 3259 AA.
AC Q14789; Q14398;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Golgi autoantigen, golgin subfamily B member 1 (Giantin) (Macrogolgin)
DE (Golgi complex-associated protein, 372-kDa) (GCP372).
GN GOLGB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94187728; PubMed=7511208;
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
RA Renz M.;
RT "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane
RT protein (giantin).";
RL Mol. Cell. Biol. 14:2564-2576 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94257116; PubMed=8198703;
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
RA Renz M.;
RT "Macrogolgin -- a new 376 kD Golgi complex outer membrane protein as
RT target of antibodies in patients with rheumatic diseases and HIV
RT infections.";
RL J. Autoimmun. 7:67-91 (1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95100974; PubMed=7802676;
RA Sohma M., Misumi Y., Fujiwara T., Nishioka M., Ikehara Y.;
RT "Molecular cloning and sequence analysis of a human 372-kDa protein
RT localized in the Golgi complex.";
RL Biochem. Biophys. Res. Commun. 205:1399-1408 (1994).
```

```
CC -!- FUNCTION: May participate in forming intercisternal cross-bridges
CC of the Golgi complex.
CC -!- SUBUNIT: Homodimer; disulfide-linked.
CC -!- SUBCELLULAR LOCATION: Golgi; rheumatoid arthritis and in the
CC DISASE: Antigen in chronic rheumatoid arthritis and in the
CC autoimmune disease Sjogren's syndrome. Usage by and for commercial
CC -!- SIMILARITY: Belongs to the golgin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X75304; CAAS3052.1; -.
DR EMBL; D25542; BAA05025.1; -.
DR FIR; A56539; A56539.
DR FIR; I52300; I52300.
DR Genew; HGNC:4429; GOLGB1.
DR MIM; 602500; -.
DR GO; GO:000139; C:Golgi membrane; TAS.
DR GO; GO:0005795; C:Golgi stack; TAS.
DR GO; GO:0016021; C:integral to membrane; TAS.
DR GO; GO:0007030; P:Golgi organization and biogenesis; TAS.
KW Golgi stack; Antigen; Coiled coil; Transmembrane.
FT DOMAIN 1 3235 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 3236 3256 POTENTIAL.
FT DOMAIN 3257 3259 LUMENAL (POTENTIAL).
FT DOMAIN 48 593 COILED COIL (POTENTIAL).
FT DOMAIN 677 1028 COILED COIL (POTENTIAL).
FT DOMAIN 1062 1245 COILED COIL (POTENTIAL).
FT DOMAIN 1301 1779 COILED COIL (POTENTIAL).
FT DOMAIN 1828 3185 COILED COIL (POTENTIAL).
FT DOMAIN 2420 2423 POLY-GLU.
FT DOMAIN 2993 2996 POLY-SER.
FT CONFLICT 1 39 MISSING (IN REF. 3).
FT CONFLICT 215 215 A -> AQLSSM (IN REF. 3).
FT CONFLICT 1765 1765 D -> G (IN REF. 3).
FT CONFLICT 2950 2950 H -> D (IN REF. 3).
SQ SEQUENCE 3259 AA; 376075 MW; 60376A20D8A178DD CRC64;

Alignment Scores:
Pred. No.: 0.0118 Length: 3259
Score: 123.50 Matches: 213
Percent Similarity: 34.91% Conservative: 155
Best Local Similarity: 20.21% Mismatches: 326
Query Match: 1.74% Indels: 363
DB: 1 Gaps: 45

US-09-602-362E-22 (1-4115) x GOBL_HUMAN (1-3259)
QY 1383 TTCAAGCCGCCATTTGAAGCAAAACTCTTCCA-----AAT 1418
Db 1158 TrpLysProGluLeuGluLysIleLeuAlaLeuGluLysGluLysGluLeuGln 1177
QY 1419 AAGCCTTTGAATGAAGATGAACAACATTTGAGCAGATCCGTTGCCACCGAATCCAA 1478
Db 1178 LysLysLeuGlnGluAlaLeuThr-SerArgLysAlaIleLeuLysLysAlaGlnGlu 1197
QY 1479 ACAAGAGACTATAAGAAATCTCTTG-----GGATTCTAGAGTCTCTGGAG 1523
Db 1197 sGluArgHisLeuArgGluLeuLysGlnGlnLysAspAspTyrAsnArgLeuGln 1217
QY 1524 ACTGTTTACAGAA-----GGATTGTTTACCAAGCTACA 1559
Db 1217 uGlnPheAspGluGlnSerLysGluAsnGluAsnIleGlyAspGlnLeuGln 1237
QY 1560 CATCAAAAAGATAGATAAAAAATGGAATAAAGAGAGTCCCTAATAAAGTGGTCTCT 1619
Db 1237 nIleGlnValArgGluSerIleAspGlyLys-----Le 1248
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 09:09:52 ; Search time 60.1449 Seconds
(without alignments)
13162.486 Million cell updates/sec

Title: US-09-602-362E-22

Perfect score: 7086
Sequence: 1 ctgctctacagcaacgac.....gagtggactccactggaaa 4115

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2.1/USPNO.spool_p/US09602362/runat_15072004_093626_22015/app_query.fasta_1.10325
-DB=pir_78 -QFWT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCTX=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -WAITRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US09602362 @CGN 1 1 248 @runat_15072004_093626_22015 -NCPU=6 -ICPU=3
-NO MMAP -LARGEOQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

pir_78:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Length | ID | Description |
|------------|-------|--------------|--------|-------------|
| C 1 | 148.5 | 2.1 | 1363 2 | T15653 |
| C 2 | 146 | 2.0 | 590 1 | S34960 |
| C 3 | 144 | 2.0 | 1642 2 | T08890 |
| C 4 | 143 | 2.0 | 445 2 | E22845 |
| C 5 | 139 | 2.0 | 1818 1 | S73852 |
| C 6 | 138.5 | 1.9 | 502 1 | I30010 |
| C 7 | 138 | 1.9 | 590 1 | Q0UTCS |
| C 8 | 136.5 | 1.9 | 443 2 | B26696 |
| C 9 | 136 | 1.9 | 1613 2 | D90129 |
| C 10 | 134 | 1.9 | 421 2 | T44511 |
| C 11 | 133.5 | 1.9 | 527 2 | S26037 |
| C 12 | 133 | 1.9 | 426 2 | T44522 |
| C 13 | 132 | 1.9 | 1979 2 | C71622 |
| C 14 | 131.5 | 1.8 | 288 2 | S36954 |

| | | | | | | |
|------|-------|-----|------|---|--------|--------------------|
| C 15 | 130.5 | 1.8 | 288 | 2 | S36953 | cytochrome-c oxida |
| C 16 | 130 | 1.8 | 528 | 2 | S26025 | NADH2 dehydrogenas |
| C 17 | 130 | 1.8 | 1163 | 2 | G97236 | ATPase involved in |
| C 18 | 130 | 1.8 | 2663 | 1 | S28261 | centromere protein |
| C 19 | 129.5 | 1.8 | 1231 | 2 | S70553 | chromosome-associa |
| C 20 | 128.5 | 1.8 | 620 | 2 | A58932 | cytochrome C-type |
| C 21 | 128.5 | 1.8 | 918 | 2 | T40030 | hypothetical prote |
| C 22 | 128.5 | 1.8 | 1410 | 1 | A57013 | early endosome ant |
| C 23 | 127 | 1.8 | 1005 | 2 | A64465 | hypothetical prote |
| C 24 | 127 | 1.8 | 2954 | 2 | T14156 | kinasin-related pr |
| C 25 | 126.5 | 1.8 | 288 | 2 | S36955 | cytochrome-c oxida |
| C 26 | 126 | 1.8 | 1187 | 2 | T20544 | hypothetical prote |
| C 27 | 125.5 | 1.8 | 2245 | 2 | T18278 | myosin heavy chain |
| C 28 | 125 | 1.8 | 1928 | 2 | S45773 | myosin heavy chain |
| C 29 | 123.5 | 1.7 | 1312 | 2 | T30845 | probable DNA repai |
| C 30 | 123.5 | 1.7 | 3259 | 1 | A58539 | giantin - human |
| C 31 | 123 | 1.7 | 313 | 2 | T11153 | NADH2 dehydrogenas |
| C 32 | 123 | 1.7 | 3225 | 2 | I52300 | giantin - human |
| C 33 | 122.5 | 1.7 | 531 | 2 | T11074 | NADH2 dehydrogenas |
| C 34 | 122.5 | 1.7 | 3187 | 2 | JCS837 | 364K Golgi complex |
| C 35 | 121.5 | 1.7 | 509 | 2 | F71526 | hypothetical prote |
| C 36 | 121 | 1.7 | 980 | 2 | E71606 | hypothetical prote |
| C 37 | 120.5 | 1.7 | 409 | 2 | S26033 | NADH2 dehydrogenas |
| C 38 | 120.5 | 1.7 | 1972 | 1 | A41604 | myosin heavy chain |
| C 39 | 118 | 1.6 | 513 | 2 | E71683 | NADH2 dehydrogenas |
| C 40 | 118 | 1.7 | 978 | 2 | A70387 | conserved hypothet |
| C 41 | 118 | 1.7 | 1173 | 2 | T43527 | sp8 protein - fies |
| C 42 | 118 | 1.7 | 1727 | 2 | T50073 | myosin-like coiled |
| C 43 | 117.5 | 1.6 | 478 | 2 | T11318 | NADH2 dehydrogenas |
| C 44 | 117.5 | 1.6 | 547 | 2 | T33437 | hypothetical prote |
| C 45 | 116.5 | 1.6 | 437 | 2 | S34959 | NADH2 dehydrogenas |

ALIGNMENTS

RESULT 1

T15653

Hypothetical protein C27D6.2 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C/Accession: T15653

R/Ding, H

submitted to the EMBL Data Library, July 1995

A/Description: The sequence of C. elegans cosmid C27D6.

A/Reference number: Z18383

A/Accession: T15653

A/Status: preliminary; translated from GB/EMBL/DBBJ

A/Molecule type: DNA

A/Residues: 1-1363 <DIN>

A/Cross-references: EMBL:U23179; NID:G726424; PID:G726426; PIDN:AAC46722.1; CESP:C27D6.2

A/Experimental source: strain Bristol N2

C/Genetics:

A/Gene: CESP:C27D6.2

A/Introns: 76/3; 149/3; 215/2; 287/3; 353/2; 434/3; 507/3; 613/3; 690/3; 785/1; 806/3; 8

Alignment Scores:

| | | | |
|------------------------|----------|---------------|------|
| Pred. No.: | 1.39e-05 | Length: | 1363 |
| Score: | 148.50 | Matches: | 200 |
| Percent Similarity: | 33.53% | Conservative: | 136 |
| Best Local Similarity: | 19.96% | Mismatches: | 325 |
| Query Match: | 2.07% | Indels: | 342 |
| DB: | 2 | Gaps: | 51 |

US-09-602-362E-22 (1-4115) x T15653 (1-1363)

| | | | | | |
|----|------|--|--------------------------|--------------------|------|
| QY | 3046 | GGTTTAGT--- | GGGCTATTCCAGTTTTCAT----- | GGCAATTCCCTTTTCAAC | 2999 |
| | | | | | |
| Db | 28 | IlePheSerThrLeuAlaValProAlaLeuPheIlePheLeuLeuLysGlnIlePhePro | 47 | | |
| QY | 2998 | ATGCATTCATGTGAAGATATATTTTCATTCATGAGTTGAGAACCTATTCAATTACT | 2939 | | |
| | | | | | |
| Db | 48 | LeuProPheHisGlyAsnIleLysPheMet | 57 | | |

```
QY 2938 TTCTACACCTTCAATTCTTATCTTGTATCTGAGAGCCTTTCAAGTGTGTTTCACTCTTA 2879
Db 58 -----LeuIleSerTyrPheIleSerAlaPheLeuPheAlaValValLeuAlaLeu 74
QY 2878 ACTCTTCT-----ATGCTCTCTCGANTCTT-----CTAATTCCT 2843
Db 75 ThrPheGlyTyrHisIleLeuValProLeuPheIleThrSerLysCysAspLeuIleIle 94
QY 2842 CCTAATTTTCTTAAATATATCGCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2783
Db 95 GlnProTyrLeuPheLysValGlyGlnLeuSerLeuThrLeuPheIleThr----- 111
QY 2782 CTACACTGCAGACTCTCTTCCCATTTAACTTTTGGTTCTCTACTGCAATTAATTTCTTT 2723
Db 112 -----LeuGlnMetIleMetProPheGlyPheSerIleGluArgIleAlaLeuArg 129
QY 2722 TGTTCTGACAGTTCTTTTTCAGTACACAAACTCTTTTTCATTCTTCCATTCTCTCTAC 2663
Db 130 MetAlaLysSerTyrGluAsnValArgThrValLeuGlyProLeuLeuIlePheValLeu 149
QY 2662 GTG-----TTACAGTG-----TCTTTTGAAGTTCC 2636
Db 150 IleGlyIleAspLeuIleLeuLeuPheThrValPheArgaspGluSerPheAsnAspSer 169
QY 2635 CTGCGCTTCAAGAGTGAAGTGTACCAAGATTTTGTATAGCTAGTGAATCTCTAATTT 2576
Db 169 ----- 169
QY 2575 TCCACTATTTATCCTTTCTTTTGTGTAGCTGGGTACACACCATCCTTCTGTGAAA 2516
Db 170 -----PheIleSerPheIleLeuIle-----ProIleThr 180
QY 2515 CAGTCCAGGAGACTTCAGAATCCAGAAATTTCTCAACCTCTTGTGTGATCTG-- 2458
Db 181 AlaGlnThr-----PheAsnSerTyrCysTrp-IleLeuLe 192
QY 2457 -----AAGGAAATCTGATCTGT-----CTCAATGTGTCTCAATCTTAAT 2417
Db 192 uTyrAlaGluLeuGlyAsnLeuLeuCysAsnCysIleIleLeuLeuValHisSerLysP 212
QY 2416 CCAAGGTTTATTTGGAACAGACTTTTCAT----- 2388
Db 212 eIys-----ThrAsn-TyrHisValValAlaProLeuPheIleThrSerMetC 228
QY 2387 -----TTCAATGAGGCTCAAGCAGATGGCTCTCGGA 2354
Db 228 ysAspLeuIleIleArgProSerLeuTyrLysValGlyAsnLeuSerLeuThrLeuPheM 248
QY 2353 GGCCTGTCTTGAAGCTTTGCATGCCA-----TCAATTCAGGCTTTTAGTGGAAATAG 2303
Db 248 et--ThrIleGlnMetIleMetProLeuGlyPheSerIleGluArgPheIleAlaLeu 267
QY 2302 AACTTTCATTCACGGAGCCTTCAGAAACCTCATTTACAGACTCTCTCTATTTCCTCA 2243
Db 267 exMetThrLysSerTyrGluAsnValArgThrPheLeuGlyProLeuLeuValPheThrL 287
QY 2242 TTATTTATCTATTCTTTTGTATGTAGCTGGGTA--AACACCATCCTCTCTGAAACAG 2184
Db 287 euIle-----GlyIleAspLeuAlaLeuLeuTyrHisV 298
QY 2183 TCTACAGAGACTCCAGA-----ATCCAGAGACTT 2154
Db 298 alphaArgaspGluLysPheGluAspSerPheIleSerPheAlaLeuValProGluThrS 318
QY 2153 TCTCATAGCTTTTGTGATTTGATGGAGTTCTCATCTGCCTCA----- 2107
Db 318 exAlaIleProPheAsnSerTyrPheTrpGluLeuLeuTyrAlaGluIleGlyAsnPheI 338
QY 2106 -----ATGTTTGTTCATCTTAATTCGAAGGTTTATTTCGACAGACTT 2064
Db 338 leCysAsnCysIlePheLeuLeuValHisSerLysPheLysAlaArgValHisGlnMetL 358
QY 2063 TTGATTTCAATGAGGCTCAAGGAGAGATGGCTCTCGGGAGGCTCTTGAAGTTTCCA 2004
```

```
Db 358 euGluThrAsn-----AspSerValCysGluLeuAlaTyrGlnLeuAlaT 373
QY 2003 TGTCTCATCAATCTAGGCTTTAGT-----TGAAATAGAACTTTC-----ATTC 1962
Db 373 yrHisProValTyrArgSerSerGlnPheTrpSerMetLeuValSerSerLeuSerIleP 393
QY 1961 CCAGTTAGCTCAGAGACCACTTTTAAACAGGAGCCCTTCTAATTTTCCAT----- 1912
Db 393 roAlaLeuIleTyrPheIleThrArgLysIlePhePheLeuHisPheHisGlyAsnLeuL 413
QY 1911 -----TATTTTATCTATTCTTTTGGCGCAGCCTTGGTAAACACAAT 1869
Db 413 ysCysLeuLeuIleValTyrPheIle----- 421
QY 1868 CCTTCTGTGAACAGCTCTACAGAGACTCCAGTATCCAGATTTTCTTATAGTCCCTTGT 1809
Db 422 -----CysAsnLeuLeu-----PheSerMetAlaLeuCysP 432
QY 1808 TTGGATTTCTGATGGGAGTTCTCATCTGCTCCATGTTTTCATTTTTCATTTCCAGGCTT 1749
Db 432 heAlaPhePheTyrGlnPhe-----LeuIleProPhePheValThrSerLysCysGlnL 450
QY 1748 ATTTGGGACAGACTTTTGATTTTCACTGGAGGCTCGAAGCAGATGGCTTCCCGAGGCTCG 1689
Db 450 euLeuIleAsnThrThrLeuPheLysTrpGly-----GlnIleCysSer--PheLeuL 467
QY 1688 CTTTGAAGATTTGCAATGCTTCAATTCAGGCTTTAGTGGATAGAAATTAATTCATCCGAG 1629
Db 467 euLeuThrSerSerMetLeuLeuProIleGlyPheSerIleGluArgPheValAlaLeuG 487
QY 1628 GTAGCCTTCAGAG-----ACCACATTTATTAGGACTCTTCTATTTCATTT 1581
Db 487 IyAsnAlaGlnLysTyrGluSerSerArgThrPheLeuGlyProValIleIlePheIleI 507
QY 1580 TTTTA-----TCTATTCTTTTGTATGTG-----TAGCCTCGTAAACAC 1542
Db 507 leIleAlaValAspPheSerIleIlePheSerValTyrLysAsnGluProPhe--ThrG 526
QY 1541 AATCCTCTCTGTAAA-----CAGTCTCCAGAGACTCTAGATCCCAAGAAATTTCTTAT- 1489
Db 526 luGlyPheTyrSerPheIleLeuValProSerThrAlaSerGlnIleAsnMetTyrP 546
QY 1488 -----AGTCCTTTTGTGGATTCGGTGGGAACACGAGATCGTCTCAAT 1446
Db 546 hePheValLeuLeuPheValLysIlePheAsnLeu-----LeuLeuAsnC 561
QY 1445 GTTTGTTTCATT--CTTCATTCAAAGGCTTATTTCGGAAGAGTTTTTGTCTTCAATGGCG 1389
Db 561 ysIleLeuLeuArgIleHisLysLysIleArgIleLysTyrTyrSerLeuSerValArgT 581
QY 1388 CTTGAAGGAGAGTGGCTTTTAGGAGGCTTCTACTTCTTATTATCTCATTTCTTTAT 1329
Db 581 yrGluMetGluGluIleLeuGlnSerSerLysPheThrPheIleIleArgPheThrHisL 601
QY 1328 ATATAGACTCAGGTATA-----CCACTTGAATCTT 1299
Db 601 euLeuPhePheGlyPheTyrValValIleLeuPheLysArgMetIleGluGlnC 621
QY 1298 GCAGAACTTCAAGACACCGAGAAATCAAGAAATATTC----- 1261
Db 621 ysGluLeuAlaTyrGlnLeuAsnAsnHisProIleTyrThrIleSerGluLeuTrpSerC 641
QY 1260 -----TTCATCTTCCCTTGTGTG----- 1243
Db 641 ysLeuValSerLeuPheAlaIleProSerLeuIleTyrPheLeuThrArgLysValMetL 661
QY 1242 -----GATCTGATGGGACCTTCATCATGGCACTTGTCTTGTAGATGATTC 1197
Db 661 euMetArgPheHisGlyAsnLeuLysValLeuLeuMetAlaTyrPheSerSerLeuLeuL 681
QY 1196 TTTTGTGGACATGCACAATCTTAGATTTCTCTTTTCAAAACCTTTAGTTTATTAGATGTTA 1137
Db 1196 TTTTGTGGACATGCACAATCTTAGATTTCTCTTTTCAAAACCTTTAGTTTATTAGATGTTA 1137
```

```

Db      691 euLeuSerThrValThrPheLeuGlyTyrIleTyrGln-----IleT 695
QY      1136 CT-----CTTCACGCATCCATCTTTACAGGG 1110
Db      695 hrSerSerGlnAlaSerAsnSerLysCysGlyLeuLeuGluSerLysIlePheLysL 715
QY      1109 TTTCTTTTCTCCCATCGCTTCTCTAGGCTTCTCTTTCTGCCCCACGTAATTTCTCAGT 1050
Db      715 euSerHisIlePro----- 719
QY      1049 GTTTCCTTTTCGGGACTCATATTTCCCTAGGTGTGTCTCTTTTCTCCCTGCGCATCTCT 990
Db      720 -----LeuLeuPheMetLeuThrCysSerLeuLeuLeuProIleGlyPheS 735
QY      989 AGGTCTTCCTTTGTCGCCCGCTAAATTTCTCAGANGTTTTCAGAGATCGTAATTTCC 930
Db      735 erIleGluArgSerIleAlaLeuArgMetAlaLysLysTyrGluHisValArgThrPheL 755
QY      929 TAGGT-----GTTTTCTCTCTGACG----- 910
Db      755 euGlyProIleLeuValPheIleLeuValArgPheThrValPheIleLeuSerValLeuA 775
QY      909 -----TTCGAACTTTCCAGATGTCTCTTTCTCCAACTT----- 877
Db      775 snTyrAspPheHisAspArgGlnGluValAsnGlnValIleIlePheValAlaAlaM 795
QY      876 -----GAATTTCTCAGATGTTCTCCACCAAGATGCAGCCTCATCAGGTGTTTT 828
Db      795 etLeuThrAsnGlnMetIleArgTyrTyrSerTyrLysLysLysMetAlaGluile----- 813
QY      827 TCACACGAGCTTCA-----GCCTGTGAGGTGTCTTCCACCAAGGTGCAGCCTCAT 777
Db      814 --AsnGlnThrLysCysAspLeuAlaPheGlnIleSerTyrHis----- 827
QY      776 CAGGTGTTTTCCACGAGCTTTCAGCGTGCAGGTGCTTTCGCCGACGGGTGCAGCTCA 717
Db      828 -----ProIleTyrArgLeuAlaGlnPheTrpThrLeuSerValSerLeuL 843
QY      716 TCAGGTTTCTCGCAGAGCTTCTCTCGATTGCTATTTTCATGATTTTTCATATTTTCGTAT 657
Db      843 euAlaValProSer-----LeuLeuTyrPheL 852
QY      656 ATATTCAATAATTTGTTANGAATGTGAGAAATCCACAGTACAGCATATGTTCTGCAGTTA 597
Db      852 euLeuLysArgValLeu---LeuLeuProPheHisGlyAsnLeuLysCysLeuLeuIleT 871
QY      596 CT 595
Db      871 hr 871

RESULT 2
S34960
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Crithidia oncopelti mitochondria
C:Species: mitochondrion Crithidia oncopelti
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Jun-2002
C:Accession: S34960
R:Maslov, D.A.; Horvath, A.; Gwang II, K.; Kolesnikov, A.A.
submitted to the EMBL Data Library, October 1990
A:Reference number: S34958
A:Accession: S34960
A:Molecule type: DNA
A:Residues: 1-590 <MAS>
A:Cross-references: EMBL:X56015; NID:gl2879; PIDN:CAA39492.1; PID:gl2882
C:Genetics:
A:Gene: NDS
A:Genome: mitochondrion
A:Genetic code: SGC6
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Alignment Scores:
Pred. No.: 2,08e-05 Length: 590
Score: 146.00 Matches: 121

```

```

Percent Similarity: 35.34% Conservative: 61
Best Local Similarity: 23.50% Mismatches: 158
Query Match: 2.03% Indels: 176
DB: 1 Gaps: 27

US-09-602-362E-22 (1-4115) x S34960 (1-590)

QY      3977 TTTATGCTGTAGGCTTCAGGAAGAACATGAATTTTCTTAAATAGTATTCTGT--- 3921
Db      76 PheTyrCysAsnGlyPheTyrLeuPheIleLeuPheLeuLeuValPheCysPhe 95
QY      3920 -----GACACAGACTTTGGTAAGATTTCAC 3894
Db      96 IleLeuPheTyrAlaPheTyrTyrMetTyrTyrAspLeuMetLeuLysArgPheAsn 115
QY      3893 ATAGGTGATGTAGACTGGCCCTCTAGCATGAGTGTGAGTAAAGTCTGTCTCTGTCTT 3834
Db      116 Ile-----PheTyrTrpPheValLeuCys 123
QY      3833 CCAAGAAGATTTCCTACTGCT-----TGTTCTCATGAGTTTCTG 3796
Db      124 MetAsnPhePheIleLeuSerTyrAspTyrLeuThrAlaTyrCysGlyTrpGluLeuLeu 143
QY      3795 TTTCTCTTCTCTTTTTCATATTGAAT-----ATACGGTGTTTTAAATCGTATTGTA 3745
Db      144 GlyLeuPheSerPhePheLeuIleSerTyrPheTyrTyrArgPhePhe---AlaLeuLys 162
QY      3744 ATATAATCTCTCTTTT-----TCTCTTTAGAGATGTGTTCATTTCTCTCAAGA 3691
Db      163 PheGlyPheLysSerPhePheIleSerLysIleGlyAspValLeuLeuLeuSer--- 181
QY      3690 AATGAATATCATTTGTTATCTGCTT-----TTGTTGTCAGCT 3655
Db      182 -----PheValMetThrPheIleSerThrGlyTyrGlyMetIleAsnPhe 196
QY      3654 TTCTATGTGCATGACTAATTTGCTGTGAAGCCACAAATTTTGTCTGTAGTTGAATAATT 3595
Db      197 TyrPheValAsn-----Phe 201
QY      3594 CTGACTAGAGAC-----TCTGCTGTTCAGGTGTTTGTTCATTAATCTTCTGCTTT 3544
Db      202 LeuCysValAspPheTyrPheThrAlaPheMetLeuPheLeuLeuIleMetCysAlaPhe 221
QY      3543 GATACATGTGTTGCGCTTC---CTTCATTTGACACTGTTTTCACGTTGCTCTTTGTGATGT 3487
Db      222 ThrLysSerThrGlnPheGlyLeuHis----- 230
QY      3486 TCTGAACCAATGTATTTCTCTTAGACATCTCTCTAATTGAGATAATTTTGTAGTTTTT 3427
Db      231 -----IleTrpLeuProAspAla-----MetGluGly 239
QY      3426 GGATTTCTTTTCAGCTCAGAAAGTGTGTGAGCAGCTCATTTGTTAATATCTACTCTCA 3367
Db      240 ProIleProValSerAlaLeuIleHisAlaAlaThrLeuValValCysGlyIleLeuLeu 259
QY      3366 CATCAACATTCATTTCTTTTGCACACAGCATCTCTCAATGTGGAAGCAGGTTCTTACTT 3307
Db      260 -ValSerPhePheTyrPheTyrCysPheAspPheTrpLeu-----AlaTyr-Phe 274
QY      3306 TTTCTTGTATG-----TCACAAATTCATGCTTGTGTACAGCAGAACCCAGTC----- 3261
Db      274 eTyrProLeuIleGlyTrpSerSerLeuIleLeuValMetMetSerLeuCysValPheTyr 294
QY      3260 -----AGGATGCTGTATTCAATTTCTGCTCTCTAG 3232
Db      294 rAsnPheAspAlaLysArgPheValAlaPheSerThrIleCysGlnIleSerPheSerMe 314
QY      3231 TTTCTTTTGTCTGTTTTTCTCAATTTAGATGAGCATTTGTTCTCAGCTATCAGAACTT 3172
Db      314 tPheCysCysLeuCysLeuAspLeuTyrValGlyCysLeuPhePheCysTyrHisMetPh 334
QY      3171 TAGCTGCCACACATATTGAGATCCCTTTTAGTAATGATTCTCTTTTTCAGTTTGTAGGTCTATC 3112

```

```
Db 334 e-----TyrLysAlaThrLeuPheIleValLeuGly--Val 345
QY 3111 TAAGTTCAGCATCTTTCTTTTAA----- 3088
Db 346 TrpIle-----HisLeuPhePheGlyLeuGlnAspValArgCysTyrPhePheIleTyrPhe 364
QY 3087 -----ATCTTAAGTCCTCAAGATTTATTTTCTCTTTTCTCGGATTTGGTGTAGTG 3037
Db 365 CysGlyCysValLeuAlaArgMetLeuLeuValPheAlaLeuLeuAsnSerCysSerLeu 384
QY 3036 TGGCTATTCCAGTTTGGCATGGCAATCTCTTTTCAACATGCATTCATTCATGTAAGATAA 2977
Db 385 TrpPheLeu-----CysGlyPheTyr----- 391
QY 2976 TTTTCATTTCATGAGTTTGAGAAACCTATTCAAAATTACTTTCTACACCTTCAATTCTTATC 2917
Db 392 -----CysLysAspLeu-----LeuLeuCysThr----- 399
QY 2916 TTGATATCGAGAGCCTTCAAGTTGTGTCTCACTTCACTTCACTTCTTATGCTGCTCTCGA 2857
Db 400 LeuMetLeuValSerPheHisPheIleLeuGluPhe---LeuPheValCysIle----- 416
QY 2856 TTCCTCTAATCTCTCTAATTTTCTTTTCTTAATATATCGCATTTCTCTCTCTCTTTCT 2797
Db 417 -----PhePheIlePhePheThrValIleTyrAsnTyrPheLeuLeuPhePheLeu 433
QY 2796 TGG-----TTTAAGTCAATCTACACCTGTCAGACTCTTGTTCCTCCATTTAATCTTTGTTCT 2743
Db 434 CysPheValPheLysCysPheCysLeuAlaAsp---CysLeuPheLeuLeuPheAspPhe 452
QY 2742 CTACTGTGATTATTTCTTTCTGTGACAGTTCTTTTTCAGTACACAAACTCTTTTTC 2683
Db 453 GluCysCysLeuIleTyrCysPheLeuGlyLeuTyrMetCysPheIleLeuIlePhePhe 472
QY 2682 ATTGTTCATTCTCTGTACGTGTTTCACAGTGTCT 2648
Db 473 ValIleAsp---PheLeuTyrIlePheValPheSer 483
RESULT 3
T08880
NM_001001: receptor-binding protein yotiao - human
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T08880
R:Lin, J.W.; Wyszynski, M.; Madhavan, R.; Sealock, R.; Kim, J.U.; Sheng, M.
J. Neurosci. 18, 2017-2027, 1998
A:Title: Yotiao, a novel protein of neuromuscular junction and brain that interacts with
A:Reference number: Z16511; MUID:98151389; PMID:9482789
A:Accession: T08880
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1642 <LIN>
A:Cross-references: EMBL:AF026245; NID:g2623067; PIDN:AB86384.1; PID:g2623068
C:Genetics:
A:Map position: 7q21-22
C:Keywords: brain; cerebral cortex; coiled coil; neuromuscular junction; skeletal muscle

Alignment Scores:
Pred. No.: 3,97e-05 Length: 1642
Score: 144.00 Matches: 175
Percent Similarity: 39.36% Conservative: 143
Best Local Similarity: 21.66% Mismatches: 252
Query Match: 2.03% Indels: 240
DB: 2 Gaps: 39

US-09-602-362E-22 (1-4115) x T08880 (1-1642)
QY 1989 CTAGAATTGATCATGCAATGCAAACTTCAAGAGGCGCTCCGAGAGCCATCTCCCTTCGAGC 2048
Db 169 LeuGluMetMet-----GluSerGluLeuAlaGlyLys 179
QY 2049 CTCCTTGAATCAAAAGTCTGTCCAAATTAACCTTGGAAATTAGAAATGAACAAACATTTG 2108
```

```
Db 180 GlnHis-GluIleGluGluLeuAsnArgGluLeuGluMetArgVal----- 195
QY 2109 AGGCAGATGAGAACTCCCATCAAAATCCAAACAAAGGACTATGAGAA---AGTCTTGGGA 2165
Db 196 -----ThrTyrGlyThrGluGlyLeuGlnLeuGlnGluPheG1 209
QY 2166 TTCTGGAGTCTCTGATAGACTGTTTCCAGAGGATGGTGTATTTTACCCAGCTACACATCAA 2225
Db 209 uAlaAlaIleLys-----GlnArgAspGlyIleIleThrGlnLeuThrAla-- 224
QY 2226 AGAATAGATAAAATAATGMAAATAGAGAGTCTCTGATAAT----- 2268
Db 225 -----AsnLeuGlnGlnAlaArgGluLysAspGluThrMetArgGluPheLe 241
QY 2269 -GAGTTTTCGAGGCTCCCTCGAAGATGAAGTTCTATTCACATAAGCCTTGAATTGA 2327
Db 241 uGluLeu-ThrGluGlnSerGlnLysLeuGlnIleGlnPheGln-----GlnLeuG 258
QY 2328 TGGCATGCAAACTTTCAAGCAGGCTCCGAGAGCCATCTGCTTCGAGCCTGCAATGAA 2387
Db 258 lAlaSerGluThrLeuArgAsnSerThrHisSerThrAlaAlaAspLeu-LeuGln 277
QY 2388 ATGAAAGTCTCTTCCAAATTAACCTTGGAAATTAAGATGAACACATTTGACAGATCA 2447
Db 278 AlaLysGln-----GlnIle-----LeuThrHisGlnGlnLeuGluGln 292
QY 2448 GATTTCCCTTCAGATCAAAACAAAGAGGTTGAGAAATTTCTGGATTCTGAAGTCTCC 2507
Db 293 AspHisLeuLeuGluAspTyrGlnLysLysLysGluAspPhe-----ThrMetGln 309
QY 2508 GTGAGTCTTTCCAGAGGATGGTGTGTACCCAGGCTACACACAAAGAAAGAGATAA 2567
Db 310 lIleSerPheLeuGlnGlu-----LysIleLysValTyrGluMetGluGlnAspLys 326
QY 2568 ATATGGAATAATAGAGATTCACTAGCTATCAAAATCTTGGTACAGTTCTATCTGTGA 2627
Db 327 LysValGluAsnSerAsn----- 332
QY 2628 AAGGCAAGGAACTTCAAAAGACACTGTG-----AACAGCTACAGGAAATG 2675
Db 333 LysGluGluIleGlnGluLysGluThrIleIleGluGluLeuAsnThr---LysIleIle 351
QY 2676 GAACAATGAAAGAGATTTTG-----TGACTGAAAGAAAGACTG 2714
Db 352 GluGluGluLysThrLeuGluLysAspLysLeuThrThrAlaAspLysLeuLeu 371
QY 2715 TCAGACAAAGAAATAATACATCACAGTAGAGAACCAAAAGTTAAATGGGAACAGAGTCTG 2774
Db 372 GlyGluGlnGlnIleValGlnLysAsnGlnIle-----LysAsnMet 388
QY 2775 CAGTGTAGATTGACTTAAACCAAGAAAGAGAGAGAAATGCCGATATATTAAAGAAA 2834
Db 389 LysLeuGluLeuThrAsnSerLysGlnLysGluArgGlnSerSerGluGluIleLysGln 408
QY 2835 AAATTAGGAGATTAGAGAAATCGAGACAGC----- 2867
Db 409 LeuMetGlyThrValGluGluLeuGlnLysArgAsnHisLysAspSerGlnPheGluThr 428
QY 2868 -----ATAGAAAGAGTTAGAAAGTGAACACAA-----CTTGAAGGCTCTC- 2908
Db 429 AspIleValGlnArgMetGluGlnGluThrGlnArgLysLeuGluGlnLeuArgAlaGlu 448
QY 2908 ----- 2908
Db 449 LeuAspGluMetTyrGlyGlnGlnIleValGlnMetLysGlnLeuIleArgGlnHis 468
QY 2909 -----AGATACAAAGATAAGAGGTGTAGAAAGTGTAGAAAGTAATTTGAATAGG 2953
Db 469 MetAlaGlnMetGluGluMetLysThrArgHisLysGlyGluMetGluAsnAlaLeuArg 488
QY 2954 TTCTCACTCATGAATGAAATTAATTTCTTACATGAATTTGCAT-----GTT 3001
Db 489 SerTyrSerAsnIleThrValAsnGluAspGlnIleLysLeuMetAsnValAlaIleAsn 508
```

```
QY 3002 GAAAAAGAAATGGCATGCAAAATCGAATAGCCACACTAAACACCAATCCAGGAAAAGG 3061
Db 3002 GAAAAAGAAATGGCATGCAAAATCGAATAGCCACACTAAACACCAATCCAGGAAAAGG 3061
QY 509 GlnLeuAsnIleLysLeuGlnAspThrAsnSer-----GlnLysGluLysLeu 524
Db 509 GlnLeuAsnIleLysLeuGlnAspThrAsnSer-----GlnLysGluLysLeu 524
QY 3062 AAAATAAATCTTTCAGAGACTTAAGATTTTAAAGAAAAGATGCTG-----AACTTA 3111
Db 3062 AAAATAAATCTTTCAGAGACTTAAGATTTTAAAGAAAAGATGCTG-----AACTTA 3111
QY 525 LysGluGluLeu---GlyLeu---IleLeuGluGluLysCysAlaLeuGlnArgGlnLeu 542
Db 525 LysGluGluLeu---GlyLeu---IleLeuGluGluLysCysAlaLeuGlnArgGlnLeu 542
QY 3112 GATGAC-----CCT 3120
Db 3112 GATGAC-----CCT 3120
QY 543 GluAspLeuValGluGluLeuSerPheSerArgGluGlnIleGlnArgAlaAaGlnThr 562
Db 543 GluAspLeuValGluGluLeuSerPheSerArgGluGlnIleGlnArgAlaAaGlnThr 562
QY 3121 AAAACTGAAGAGGAATCATTA------AAAGGATCTCAATATGCGGCAG 3168
Db 3121 AAAACTGAAGAGGAATCATTA------AAAGGATCTCAATATGCGGCAG 3168
QY 563 IleAlaGluGlnGluSerLysLeuAsnGluAlaHisLysSerLeuSerThrValGluAsp 582
Db 563 IleAlaGluGlnGluSerLysLeuAsnGluAlaHisLysSerLeuSerThrValGluAsp 582
QY 3169 CTAAAGTTCTGATAGCTGAGAAACAATGCTCAATCTAAATGAGAAAACACAGACAAAGA 3228
Db 3169 CTAAAGTTCTGATAGCTGAGAAACAATGCTCAATCTAAATGAGAAAACACAGACAAAGA 3228
QY 583 LeuLys-----AlaGluIleValSerAlaSerGluSerArgLysGluLeuGluLeu 599
Db 583 LeuLys-----AlaGluIleValSerAlaSerGluSerArgLysGluLeuGluLeu 599
QY 3229 AAATAGAGGAGAAAT-----GAATACACCATCTGACTGCTGCTGCTGACAGCCATG 3285
Db 3229 AAATAGAGGAGAAAT-----GAATACACCATCTGACTGCTGCTGCTGACAGCCATG 3285
QY 600 LysHisGluAlaGluValThrAsnTyrLysLysLeuGluMetLeuGluLysGluLys 619
Db 600 LysHisGluAlaGluValThrAsnTyrLysLysLeuGluMetLeuGluLysGluLys 619
QY 3286 ATCAATGTGCATCAAGAAAAGTAAAGAACTGCTCCACATTGAGGAGATGCTGTTG 3345
Db 3286 ATCAATGTGCATCAAGAAAAGTAAAGAACTGCTCCACATTGAGGAGATGCTGTTG 3345
QY 619 sAsnAlaValLeuAspArgMetAlaGluSer----- 629
Db 619 sAsnAlaValLeuAspArgMetAlaGluSer----- 629
QY 3346 CAAAGAAAATGAATGTGTAGTGTAGTACGATATTAACAAATAGTCTCCATCACCCT 3405
Db 3346 CAAAGAAAATGAATGTGTAGTGTAGTACGATATTAACAAATAGTCTCCATCACCCT 3405
QY 630 -GlnGluAlaGluLeuGluArgLeuArgThrGlnLeuLeuPheSerHisGluGluLeu 649
Db 630 -GlnGluAlaGluLeuGluArgLeuArgThrGlnLeuLeuPheSerHisGluGluLeu 649
QY 3406 TTCTGAGCTCAAGGAATCCAAACCTTAAATATCTCAATTAGCAGGAGATGCTAA 3465
Db 3406 TTCTGAGCTCAAGGAATCCAAACCTTAAATATCTCAATTAGCAGGAGATGCTAA 3465
QY 649 uSerLysLeuLysGluAspLeuGluIleGluHisArgIleAsn-----IleGluLys 666
Db 649 uSerLysLeuLysGluAspLeuGluIleGluHisArgIleAsn-----IleGluLys 666
QY 3466 GAGAAATACATGTTTTCAGAACATCAACAGAGACAAACCTGAAACAGTGTCAATGAAG 3525
Db 3466 GAGAAATACATGTTTTCAGAACATCAACAGAGACAAACCTGAAACAGTGTCAATGAAG 3525
QY 666 sLeuLysAspAsnLeuGluLysHisLysGln-GlnIleAspGlyLeuGlnAsnGluM 686
Db 666 sLeuLysAspAsnLeuGluLysHisLysGln-GlnIleAspGlyLeuGlnAsnGluM 686
QY 3526 GAAGCGAACACATG-----TATCAACAGACAGATAATGGAACACACCTG 3573
Db 3526 GAAGCGAACACATG-----TATCAACAGACAGATAATGGAACACACCTG 3573
QY 686 etSerGlnLysIleGluThrMetGlnPheGluLysAspAsnLeuIle-ThrLysGlnAsn 705
Db 686 etSerGlnLysIleGluThrMetGlnPheGluLysAspAsnLeuIle-ThrLysGlnAsn 705
QY 3574 AACAGCAGCTCTAGTCAGAAATTTTCAACTACAAGCAAAAATTTGTG-----GCT 3627
Db 3574 AACAGCAGCTCTAGTCAGAAATTTTCAACTACAAGCAAAAATTTGTG-----GCT 3627
QY 706 GlnLeuIleLeuGluIleSerLysLeuLysAspLeuGlnGlnSerLeuValAsnSerLys 725
Db 706 GlnLeuIleLeuGluIleSerLysLeuLysAspLeuGlnGlnSerLeuValAsnSerLys 725
QY 3628 TCACAGCAATAGTCATGACATAGAAAGCTGACAAACAAAGACAGATAACATGATTC 3687
Db 3628 TCACAGCAATAGTCATGACATAGAAAGCTGACAAACAAAGACAGATAACATGATTC 3687
QY 726 SerGluGluMetThrLeuGlnIleAsnGlu---LeuGlnLysGluIleGluIleLeu--- 743
Db 726 SerGluGluMetThrLeuGlnIleAsnGlu---LeuGlnLysGluIleGluIleLeu--- 743
QY 3688 ATTTCTTGAGAGGAAATGCAACACATCTCTTAAAGAGAAAAAGAGAGAGATATTAATTC 3747
Db 3688 ATTTCTTGAGAGGAAATGCAACACATCTCTTAAAGAGAAAAAGAGAGAGATATTAATTC 3747
QY 743 ----- 743
Db 743 ----- 743
QY 3748 AATACCAATTTAAACCGTATATTCAATATGAAAGAGAAAGAGAAACAACTCATG 3807
Db 3748 AATACCAATTTAAACCGTATATTCAATATGAAAGAGAAAGAGAAACAACTCATG 3807
QY 744 -----ArglnGlnGluLysGluLys----- 750
Db 744 -----ArglnGlnGluLysGluLys----- 750
QY 3808 AGAACAGCAGTAAACTCTTTGGAGAAACACACAGCAGACTTTACTCACACTCATG 3867
Db 3808 AGAACAGCAGTAAACTCTTTGGAGAAACACACAGCAGACTTTACTCACACTCATG 3867
QY 751 -----GlyThrLeuGlnGluValGlnGluLeuGln 761
Db 751 -----GlyThrLeuGlnGluValGlnGluLeuGln 761
QY 3868 CTAGAGGCCAGCTAGCATCATCCTATGTTGAAAATCTTACCAAGAGTCTGTGTCCAGCAAT 3927
Db 3868 CTAGAGGCCAGCTAGCATCATCCTATGTTGAAAATCTTACCAAGAGTCTGTGTCCAGCAAT 3927
QY 762 LeuLysThrGluLeuLeuGluLysGlnMetLys-----GluLysGluAsn 776
Db 762 LeuLysThrGluLeuLeuGluLysGlnMetLys-----GluLysGluAsn 776
QY 3928 ACTATTTTAGAGAAATTCATGTTTCTTCTGAGCCTACAGCATATAAATAAGTGTGA 3987
Db 3928 ACTATTTTAGAGAAATTCATGTTTCTTCTGAGCCTACAGCATATAAATAAGTGTGA 3987
QY 777 AspLeuGlnGlu--LysPheAlaGlnLeuGluAlaGluAsnSerIleLeuLysAspGluL 796
Db 777 AspLeuGlnGlu--LysPheAlaGlnLeuGluAlaGluAsnSerIleLeuLysAspGluL 796
```

```
QY 3988 AGATTACTTGTTCGAATTCATTAAGCTGCACAGGATTCCTCATCTACCTGATGTGCAGCA 4047
Db 3988 AGATTACTTGTTCGAATTCATTAAGCTGCACAGGATTCCTCATCTACCTGATGTGCAGCA 4047
QY 796 ysLysThrLeuGluAspMetLeuLysIleHis---ThrProValSer---GlnGluGluA 814
Db 796 ysLysThrLeuGluAspMetLeuLysIleHis---ThrProValSer---GlnGluGluA 814
QY 4048 GACTCATTCATCAATCAACCAAGATCCGCTCTGCACTCCAGCTAGTCACAGAGTGACTCCA 4107
Db 4048 GACTCATTCATCAATCAACCAAGATCCGCTCTGCACTCCAGCTAGTCACAGAGTGACTCCA 4107
QY 814 rGLeuIlePheLeuAspSerIleLysSerLysLys-----AspSerV 829
Db 814 rGLeuIlePheLeuAspSerIleLysSerLysLys-----AspSerV 829
QY 4108 CCTGGAAA 4115
Db 4108 CCTGGAAA 4115
QY 829 alTrpGlu 831
Db 829 alTrpGlu 831
RESULT 4
E22845
Hypothetical protein 4 - Trypanosoma brucei mitochondrion
C/Species: mitochondrion Trypanosoma brucei
C/Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 11-May-2000
C/Accession: E22845
R/Hensgens, L.A.M.; Brakenhoff, J.; De Vries, B.F.; Sloof, P.; Tromp, M.C.; Van Boom, J.
Nucleic Acids Res. 12, 7327-7344, 1984
A/Title: The sequence of the gene for cytochrome c oxidase subunit I, a frameshift contain-
ing a maxi-circle DNA.
A/Reference number: A93537; MUID:85037915; PMID:6093040
A/Accession: E22845
A/Molecule type: DNA
A/Residues: 1-445 <HEN>
C/Genetics:
A/Genome: mitochondrion
A/Genetic code: SG6
C/Suprafamily: hypothetical protein 1 (CYB-COI intergenic region)
C/Keywords: mitochondrion
Alignment Scores:
Pred. No.: 3-87e-05 Length: 445
Score: 143.00 Matches: 95
Percent Similarity: 40.22% Conservative: 88
Best Local Similarity: 20.88% Mismatches: 162
Query Match: 1.99% Indels: 110
DB: 2 Gaps: 19
US-09-602-362E-22 (1-4115) x E22845 (1-445)
```

```
QY 3836 TCTCCAAAG-----TTTCTAGTGTGTCTCATGAGTTTCTGTTCTCTTCTCTCT 3783
Db 3836 TCTCCAAAG-----TTTCTAGTGTGTCTCATGAGTTTCTGTTCTCTTCTCTCT 3783
QY 1 ThrProLysSerLeuPheLeuTyrIleIle---HisIlePheIleLeuIleIle 20
Db 1 ThrProLysSerLeuPheLeuTyrIleIle---HisIlePheIleLeuIleIle 20
QY 3782 TT---TTCATATGAATATACGGTTTTTAAATGGTATTGTAATTAATATCTCTCTTTT 3726
Db 3782 TT---TTCATATGAATATACGGTTTTTAAATGGTATTGTAATTAATATCTCTCTTTT 3726
QY 20 yrSerPheIleIleLeuCysAspTyrThrThrLeuThrLeuSerPheAspLeuLeu 40
Db 20 yrSerPheIleIleLeuCysAspTyrThrThrLeuThrLeuSerPheAspLeuLeu 40
QY 3725 CTCCTTTAGGAGATGTGTGCAATTCCTCTCAAGAAATGAATATCATTTGTTCTGCTTT 3666
Db 3725 CTCCTTTAGGAGATGTGTGCAATTCCTCTCAAGAAATGAATATCATTTGTTCTGCTTT 3666
QY 40 rPLeuIleAsnLeuPheTrpIleThrLeuLeuAspSerTyrIleCysPheIlePheI 60
Db 40 rPLeuIleAsnLeuPheTrpIleThrLeuLeuAspSerTyrIleCysPheIlePheI 60
QY 3665 TGTGTCAGCTTTCATGTGCATGACTAATTCCTGTGAAGCCACAAATTTTCTGTTAGT 3606
Db 3665 TGTGTCAGCTTTCATGTGCATGACTAATTCCTGTGAAGCCACAAATTTTCTGTTAGT 3606
QY 60 leuLeuPheLeuPhe---CysPhe-----ThrLeuPheCys--- 72
Db 60 leuLeuPheLeuPhe---CysPhe-----ThrLeuPheCys--- 72
QY 3605 TGAATAATTTCTGACTAGAGACTGCTGTTGTTGAGTGTGTTGTTCCATTCTTCTGTT 3546
Db 3605 TGAATAATTTCTGACTAGAGACTGCTGTTGTTGAGTGTGTTGTTCCATTCTTCTGTT 3546
QY 73 -----PheLeuSerPheAspThrArgPheLeuPheIleIleIle 85
Db 73 -----PheLeuSerPheAspThrArgPheLeuPheIleIleIle 85
QY 3545 TTGATACATGTTGCTTCCTTCATTTGACACTGTTTTCACGTTGCTCTTCTGTTGATGT 3486
Db 3545 TTGATACATGTTGCTTCCTTCATTTGACACTGTTTTCACGTTGCTCTTCTGTTGATGT 3486
QY 86 lleIleGlnTyrIleIleIlePheLeuPheIlePhe----- 97
Db 86 lleIleGlnTyrIleIleIlePheLeuPheIlePhe----- 97
QY 3485 CTGAACCAATGTAATTTCTTAGACATCTCTGCTAAATTTGAGATAATTTTAGTTTG 3426
Db 3485 CTGAACCAATGTAATTTCTTAGACATCTCTGCTAAATTTGAGATAATTTTAGTTTG 3426
QY 98 -----IleAsnHisIleIleIleIleSerIleLeuPheGluIlePh 111
Db 98 -----IleAsnHisIleIleIleIleSerIleLeuPheGluIlePh 111
QY 3425 GATTTCTTTGAGCTCAGAAAGTGTGTGAGAGACTCATTTGTTAATATCGTACTCTAC 3366
Db 3425 GATTTCTTTGAGCTCAGAAAGTGTGTGAGAGACTCATTTGTTAATATCGTACTCTAC 3366
QY 111 e---SerLeuLeuPheLeuLeuMetSerArgPheGlyTyrLysIleLeuVa 130
Db 111 e---SerLeuLeuPheLeuLeuMetSerArgPheGlyTyrLysIleLeuVa 130
```

```
QY 3365 ATCAACATTCATTTTCCTTTGCAACAGCATCTCTCAATGCGGAGCAGGTTCTTACTTT 3306
Db 130 lleuTrpTyTyTyTyMetLeuAsnLeuIleAsnPhelIleLeuLeuPheIleLeuLeuTy 150
QY 3305 TTCCTTGATGTCACAAATTCATGCTTGTCAGCGAAGCCAGTCAGGATGTTATCA 3246
Db 150 r-----PheMe 152
QY 3245 ATTTCGCTCTAGTTTCCTTTGCTGCTGTTTTCTCTCAATTTAGAAATGAGCATGTTTCTC 3186
Db 152 tIleLeuAsnTyTyCysPhePheLeuCys-----AspPheCysPheLe 166
QY 3185 AGCTATCAGAATTTAGCTGCCACATATG-----AGATCCCTTTTAGTAATGAT 3135
Db 166 uValPheAspGluTrpLeuGlyIleLeuCysLeuPheTyTyThrLeuLeuIleLeu-P 186
QY 3134 TC-----CTTTAGTTTGGGTCATCTAAGTTCAGCATCTTTCTTTTAAATCTTAA 3081
Db 186 heLeuLeuTyTyIleAlaPheLeuIleLeuPheMetGluGlnLeuTyTyIleArgLeu--- 204
QY 3080 GTCCTCAAGATTTATTTCTTTTCCGCGATTGGTGTGTTAGTGTGCTATCCAGTTT 3021
Db 205 -----GlyValPheIlePheIleTyMet-----LeuThrPheTyTyIleLeuPheC 220
QY 3020 GCATGGCAATTCCTTTTCAACATGCAATTTTCATGTAAGATAATTTTCATTTTCATGAGT 2961
Db 220 ysPheIleLeuIleLeuLeuIleSerPheIle-----TyPheTyTyIleL 236
QY 2960 TGAGAAACCTATTCAAAATCTTCTACACCTTCAATCTTATCTTGTACTGAGACCT 2901
Db 236 euPheIleLeuLeuLeuPheGlnSerCysThrCysValLeuIleCysLeuAsnSerP 256
QY 2900 TTCAGTGTGTTTCACTTCTAACTCTTCTATGCTGCTCTCGATCTTCTTCTTAATCTTCC 2841
Db 256 heAlaIleVal--SerLeuLeuPheValLeuSerValAsnAsnPheCysPheLeuPheL 275
QY 2840 TAAATTTTCTTAAATATATATGCGAATTCCTCTCTCTCTTCTTCTTCTTAAAGTCAATCT 2781
Db 275 euIlePheIleSerThrLysAsnTyTyIlePheTyTyLeuTyTyLeuAsnPheHisLeuIle 295
QY 2780 ACACGACAGATCTTGTTCCTTAACTTTTGTGTTCTCTACTGATGATTTATTTCTTTTG 2721
Db 295 yzSerIle-SerLeuVal-----LeuLeuIleIleIleTyTyTyPhePheIleIle 311
QY 2720 TTCTGACAGTCTTTTTCAGTACACAAACTCTTTTTCATTTGTTCCATTTTCTGTTACGT 2661
Db 312 TyzAsnIlePheAspPheLysTyzAsnGluAsnTyTyPheLeuIleAsnPhe----- 328
QY 2660 GTTCACAGTCTTTTTCAGTTCCTTTCCTTCCCTTCCACAGATGAATGATTTT 2601
Db 329 -----IlePhePheSerPhe-----PheAsnAsnPhe 337
QY 2600 GATAGGCTAGTGAATCTTCTA----- 2580
Db 338 -----LeuIleSerLeuLeuLeuAlaCysLeuPheLeuCysIleGlyAlaIleProIle 355
QY 2579 ATTTTCCATTTTATCTTCTTCTTTTGTGTG 2547
Db 356 ValPheGlyPhePheIleLysValPheCysLeu 366

RESULT 5
S73852
hypothetical protein MG218 homolog F10_orf1818 - Mycoplasma pneumoniae (strain ATCC 2934
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: S73852
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105985; PMID:8948633
A:Accession: S73852
```

```
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1818 <HIM>
A:Cross-references: EMBL:AE000051; GB:U00089; NID:G1674211; PID:AA896174.1; PID:G1674222
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: Mycoplasma genitalium hypothetical protein MG218

Alignment Scores:
Pred. No.: 0.000125 Length: 1818
Score: 139.00 Matches: 236
Percent Similarity: 35.48% Conservative: 200
Best Local Similarity: 19.20% Mismatches: 384
Query Match: 1.96% Indels: 409
DB: Gaps: 56

US-09-602-362E-22 (1-4115) x S73852 (1-1818)
QY 901 AAGTTGGAAGCTCAGCAGAAACACCTAGGAAATACGATCTCTCCAAAAAACAATCTGA 960
Db 506 LysLysGlnLysGlnAspGlnLysGluAsnAspLeuLeuIlePheGluLysGlnLeuArg 525
QY 961 GAAATTTACGGCCACCAAGGAGACCTAGAGATCGCAGGAGAAAGAACACACAC 1020
Db 526 GlnTyT-----GlnAlaAspPheGluAsnGluIleGluGluLysGlnAsnGlu 541
QY 1021 CTAGGGAATATAGTCTCCGAAAGAAACACTGAGAAATTACGTGGGAGAAAGAAAGA 1080
Db 542 Leu-----PheAlaSerGlnLysSerLeuGlnLysSerPheThrGlnLeuLysAsnLys 559
QY 1081 CCTAGGAGACCCATCGGAGAGAAAGAAACCTGTAAAGATGGAT-----GCGTGCAG 1134
Db 560 GluAlaGluLeuAsn---GlnLysAlaGlnLysIleAlaGluAsPTrPALAHISLeuLys 578
QY 1135 AGTAACATCTAATAAACTAAAGTTTGAAGAGAAATCTAAGATGTTGCATGT----- 1188
Db 579 GlnAsn-----LysHisHisAlaAspLeuLeuIlePheLeu 591
QY 1189 -----CCAAAAAGAAATCATCTACAAAGCAAGTCCTCATGATCAGAGTCCCATCAGATCAA 1245
Db 592 GluGlyGluPheAsnHisLeuGlnGlnLysHis-----LysLeuLeuGluAlaArg 609
QY 1246 ACAAGGGAAGATGAAGAAATATTCTTTGATTCTCGGGTCTCTTTGAAGTTCGCAAGATTC 1305
Db 610 ThrGlnPheAspAsnArgValSerLeuLeu-----SerAlaArgPhe 623
QY 1306 AAGTGTATACCTGAGTCTATATATAAAAGTAATGAGATAAATAAGAGTAGAAAGCCT 1365
Db 623 ----- 623
QY 1366 CCTAAAGCCATCTGCCTTCAGCCCATGAAAGCAAACTCTTTCCAAATAAGCCTT 1425
Db 624 -----LysGlnLysGlnAlaGluLeu----- 630
QY 1426 TGAATGAAGAATGAACAAACATTTGAAGCAGATCCGTGTTCCACCGAATCCAAAGAGG 1485
Db 631 ---ValLysGlnLysGlnSerLeuGlnLeuThrAlaAlaPheAsnLysGluGlnGlu 649
QY 1486 ACTATAAGAAATTTCTGGATCTCTAGAGTCTCTGGAGACTGTTTACAGAG----- 1537
Db 650 AlaVal-GluArgAspTrpLysAspArgLeuAlaAsnLeuGluLysGlnLysGluMetLe 669
QY 1538 ----GATTGTGTTTACCACAA---GGCTACACATCAAAAAAGAAATAGATAAAAAATCGAAAA 1590
Db 669 uGlyAspLysValHisGlnPheAspGluAsnSerLeuAsnIleSerLysLysLeuAlaG 689
QY 1591 TAGAGAGTCCCTTAATAAGTCG----- 1613
Db 689 uArg-GluLeuAlaIleLysPheLysGluLysGluLeuGluAlaGlnLysGlnLeuS 709
QY 1614 ----TCTTCTCAAGGCTACCTCGCAATCAAAATTT---CTATTCCCACTAAAGCCTAGAAT 1665
Db 1614 ----TCTTCTCAAGGCTACCTCGCAATCAAAATTT---CTATTCCCACTAAAGCCTAGAAT 1665
```

Db 709 erLeuAspAsnAsnAlaGlyLeuLysLeuGlnLeuAspLysLeuSerGluSerL 729
QY 1666 TCAAGACATGCAAACTTTCAAAGCGAGCTCCGGGAAG----- 1703
Db 729 euLysThrGluArgLeuGluLeuGluAlaSerLysGluArgLleLeuAspPheTyrAspG 749
QY 1704 -----CCATCTGCTT 1713
Db 749 luSerSerArgArgIleAlaAspTyrGluSerAspLeuGlnAlaArgLeuAlaGluValL 769
QY 1714 CGAGCTCCACTGAAATCAAAGTCTGTCCCAAATAAG----- 1751
Db 769 ysThrLeuGluLysAsnGlnGlnThrAlaAlaLysSerGluArgGluLeuLysValA 789
QY 1752 -----CCTTGAATCAAAATGAATAACATGAGCAGCATGAGAACTCCCATCAGAA 1803
Db 789 laLeuGluLysLeuAsnGlnAlaLysLysAlaPheLeuGlnIleArgLysGlnLeuL 809
QY 1804 TCCAAACAAAGCACTATAAGAAATTCGGGATACTGGAGTCTCTGTAGACTGTTTCACA 1863
Db 809 euGluIleAlaSerValLysGlnLeuAla----- 819
QY 1864 GAAGATTTGTTTACCAAGCTCGCTCAAAAGAAATAGATAAAATAATGGAATAATAG 1923
Db 820 -----GlnLysAlaAsnLeuLeuLysAsnGln-----G 829
QY 1924 AAGGTCCTCTTAAAGTGTCTTCTGAGGCTAACTGGGAATGAAAGTCTTATTCRACT 1983
Db 829 InAlaGluLeuAspLysGlnThrGlu----- 837
QY 1984 AAAGCTAGATTTGATGACATGCAAACTTCAAAGCAGGCTCCCGAGAGCATCTGCCTT 2043
Db 838 -----GluLeuGluAlaAlaPheLeuGluGlnAspThrAspLysLysGluLeuGluL 855
QY 2044 CGAGCTCCATTTGAAATCAAAAGTCTGTCCAAATAAACCTTGAATTAAGAATGAACA- 2102
Db 855 ysAlaLeuHis-----SerValLysSerLysGlnG 865
QY 2103 --ACATTGAGCAGATGAGAACTCCCATCAATCCAAACAAAGCACTATGAGAAAGTTCT 2160
Db 865 luLeuLeuGluArgGluArgSerPheLeuLeuGlnLysGlnArgGluPheAlaGluHisV 895
QY 2161 TG-----GGATTCGAGTCTGTAGACTGTTTCCAGAAAGGATGTTTACCAGGCTAC 2217
Db 885 alAlaGlyPheLysArg-GlnValHis-----Phe 894
QY 2218 ACATCAAAAGAAATAGATAAAATAATGGAATAAGAGAGTCTCTGATAATGAGGTTTC 2277
Db 895 LysThrThrGlnMetGlnArgLeuSerGluPheAsnLysGlnGlnSerGluGlnIle 914
QY 2278 TGAAGGCTCCCTGAGAATGAAAGTCTTATTC----- 2308
Db 915 LysArgGluThrGlu-LeuLysIleAlaPheAlaAspLeuLysLysAspTyrGlnLeuPh 934
QY 2309 -----CACTAAGCCCTTGAAT 2325
Db 934 eGluLeuGlnLysAsnGlnPheGlnGlnIleGluGlnLysHisLysGluLeuGluLe 954
QY 2326 GATGGCTGCAAACTTTCA-----AAGCAGGCTCCCGAGAGCCATCTGCTTCGAGCCT 2379
Db 954 uLeuAlaGlnLysGlnAlaGluLeuLysGlnGluLeuGluGlnLysAlaThrAlaLeuAl 974
QY 2380 GCATTGAAATGAAGTCTGTTCATAATAACCTTGAATTAAGATGAACAACATTGAG 2439
Db 974 aserGlnAspGlnAspThrVal-GlnAlaLysLeuAspLeuAlaArgGlnGlnHis----- 992
QY 2440 ACAGATCAGATTTCCCTTCAGAAATCAAAACAAAGAGGTTGAGAAATTCGGGATTCG 2499
Db 993 -----GluLeuGluLeuArgGlnAsnAlaPheAsnGlnAla 1005
QY 2500 AAGTCTCCGTGGACTGTTTCAGAGAGGATGGTGTGTACCCAGGCTACACACAAAGAA 2559
Db 1005 erLeuSerLeuAsn-----LysGlnA 1012

QY 2560 AGGATAAAATAGTGGAAAAATTAGAGATTCACTAGCCTATCAAAATCTTTGGTACAGTTCA 2619
Db 1012 rgGluGlnLeuThrAsnGln-----ValLysValLeu----- 1022
QY 2620 TCTTGTGAAGCAAGGAACTTCAAAAAGACACACTGTGAACAGCTACGAGAAATGGAAC 2679
Db 1023 -----HisGlyGluLeuLysLys-----ArgHisGluLysLeuThrL 1035
QY 2680 AATGAAAAAGAGTTTGTGTACTGAAAAAGAACTGTCAAGACAAAAGAAATAAATCACAG 2739
Db 1035 euLysAspArgLeuLeu-----AlaGluLysGluLysAspGlnHisLysLysAspAlaGluI 1054
QY 2740 TAGAGAACCAAAAGTTAAATGGGAACACAGAGTCTGCAGTGTAGATTGACTTAAACCAAGA 2799
Db 1054 le-----AsnGlnArgPheLysGlnPheGluAsnGluTyrAlaAspPheAspGlnAlaLysL 1073
QY 2800 AAAGA-----GAAGAGAGAAATCGCGATA 2823
Db 1073 ysArgGluLeuGlnGluLeuAsnGlnIleArgArgAsnLeuGluGlnSerAsnAlaSerL 1093
QY 2824 TATTAAGAAAAA-----TTAGGAAGAATTAGAAATCGAGAGCA 2865
Db 1093 euLeuLysLysArgAsnGlnLeuThrLeuAspPheAlaLeuLeuArgLysValGlnHisA 1113
QY 2866 GCATAGAA-----AGAGTTAGAGTGAACACCACTTGAAGGCTCTCAGATACAAGAT 2919
Db 1113 snThrGlnThrAsnArgValGlnLeuAsnThrGln-IleLysGluPheLeuLeuGluLys 1132
QY 2920 AAGAATTGAAGGTAGAAAGTAATTGAATAGTTCCT----- 2958
Db 1133 LysAsn-----PheGlnLysAlaSerAspGluAlaAlaLeuGlnLys 1146
QY 2959 -----CAACTCATGAATGAANA 2976
Db 1147 AlaLeuLeuIleLysArgLeuArgSerPheAlaSerLysLeuGlnLeuGlnArgGluAla 1166
QY 2977 TTATTCCTACATGAAATTCATGTTTGAAGAAAGAAATTCATTCATGCAAACTGGAATAGCCA 3036
Db 1167 LeuAlaIleGlnLysLeuGluPheAspLysArgAspGluGlnGlnLys----- 1182
QY 3037 CACTAAACCAATCCAGGAAAGGAAATAAATCTTTGAGGACTTAAGATTTTAAAGAA 3096
Db 1183 -----SerGluIleAsnAsnAlaLysLeuGlnLeuGluGlnPheLysLeu 1197
QY 3097 AAGATGCTGAATAGATGACCTTAAACTGAAAGGAAATCAATTAAGAGGATCTCA 3156
Db 1198 GluLysGlnAsnPheAspGluAlaLys-----GlnLysGlnLeuIleGluPheLysAsp 1215
QY 3157 ATATGTGGCAGCTA-----AAGTTCTGTATGCTGAGA 3189
Db 1216 GlnCysGlnArgLeuAspValGluLysArgLeuLeuLysGlnLysLeuValGlnLeuLys 1235
QY 3190 AACATGTCTCATCTTAATTTGAGG-----AAAAACAGACAAAGAAACTAGAGCGCAAAAT 3246
Db 1236 AsnLeuSerLysSerTyrLeuThrTyrLysAsnArg-AlaAspLeuSerGlnGlnLeu 1255
QY 3247 GAATACACCATCTGACTGGCTTCGCTGTACAGCCATCATCAATTGTGCATCAAGAAA 3306
Db 1255 u-----GlnHisLys 1258
QY 3307 AAGTAAGAACCTGCTTCCACATTTAGGAGATGCTGTTTGCAGAAAGAAATGAATGTTGATG 3366
Db 1258 styAlaAsnLeuLeu-GluLeuLysGluLysLeuGlnThrAlaLysArgAlaLeuAspL 1278
QY 3367 TCAGAGTACGATTAACAATGAGTGCTCCAT-----CAACCTTTCTGACTCA 3417
Db 1278 ysLysHisArg-----AlaIleTyrGlyLysMetAlaGlnPheValSerGluL 1294
QY 3418 AGGAATCCAAACCTCAAAATATCTCAAT-----ACGAGA 3456
Db 1294 euArgGlnGluLysLysGlnLeuSerAlaGlnLysGlnValAspAspLysSerArgL 1314


```
QY 3099 CTTTTCCTTAA-----ATCTTAAGTCTCTCAAGATTATTTCTTTTC 3055
    |||||
Db 371 TyrPheThrTyrPheCysGlyCysileLeuAlaargMetLeuLeuilePheAlaile 390
QY 3054 CTGGATTGCTTTAGTGTGCTATTTCAGTTTTCATGGCAATTCCTTTTCAACATGC 2995
    |||||
Db 391 LeuAsnSerCysSerLeuTrpPheLeu-----Cys 400
QY 2994 AATTTCATGTAAGAATAATTTTCATTTTCATGAGTTGAGAAACCTATTCAAAATTAATTCT 2935
    |||||
Db 401 GlyPheTyr-----CysLysAspLeuLeu----- 408
QY 2934 ACACCTCAATCTTATCTGATCTGAGACCTTTCAAGTTGTGTTCTACTTCACTC 2875
    |||||
Db 409 -----LeuCysLeuLeuMetLeuThrSerPhePheileLeuLeuGluPhe-----Leu 424
QY 2874 TTCTATGCTGCTCTCGATTCTTCTAAATCTTCCTCAATTTTCTTTAAATATATCGGCAT 2815
    |||||
Db 425 CysValCysLeu-----PhePheilePhePheThrValleTyrAsnTyr 439
QY 2814 TTCTCTCTCTCTTTCTTTCTG-----TTTAAGTCAATCTACACTGCAGACTCTTTGTCC 2761
    |||||
Db 440 PheLeuLeuPhePheLeuCysPheValPheLysCysPheCysLeuValAspThrLeuPhe 459
QY 2760 CATTTA---ACTTTTGGTCTCTACTGATGATTTATTTCTTTTCTGACAGTTCTTTT 2704
    |||||
Db 460 LeuLeuPheAspPheGluCysCysLeuValTyrCysThrPheCysLeuTyrMetCysPhe 479
QY 2703 CAGTACACAAACCTTTTTCATGTTCCATTTTCTCTGTAG-----TGTTTCACAG 2653
    |||||
Db 480 -----ValLeuLeuPhePheValLeuAspPheLeuTyrValPheilePheSerSerTyr 497
QY 2652 TGTCTTTT 2644
    |||||
Db 498 CysLeuPhe 500

RESULT 7
CQUITS
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Trypanosoma brucei mitochondrion
C:Species: mitochondrion Trypanosoma brucei
C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 03-Jun-2002
C:Accession: A04519; D22845
R:Hensgens, L.A.M.; Brakenhoff, J.; De Vries, B.F.; Sloof, P.; Tromp, M.C.; Van Boom, J.
Nucleic Acids Res. 12, 7327-7344, 1984
A:Title: The sequence of the gene for cytochrome c oxidase subunit I, a frameshift contain-
ondrial maxi-circle DNA
A:Reference number: A93537; MUID:85037915; PMID:6093040
A:Accession: A04519
A:Molecule type: DNA
A:Residues: 1-590 <HEN>
A:Cross-references: GB:M94286; NID:g343546
A:Note: this translation is not annotated in GenBank entry TRBKPGEN, release 109.0
C:Comment: The DNA sequence is from a segment of the 20-kb maxicircle, which is believed
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC6
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Alignment Scores:
Pred. No.: 0.000126 Length: 590
Score: 138.00 Matches: 107
Percent Similarity: 34.08% Conservative: 75
Best Local Similarity: 20.04% Mismatches: 162
Query Match: 1.92% Indels: 190
DB: 1 Gaps: 26

US-09-602-362E-22 (1-4115) x QQUITS (1-590)

QY 3983 ACTTTATTTATGCTGAGCTTCAGGAGAAACATGAATTTCTTTCTAAATAGTATTTC 3924
    |||||
Db 194 ThrThrPheTyrPheLeuAsnPhePheCysMetAspTyrTyrTrpIleGluPheSerIle 213
```

```
QY 3923 TGT-----GACACAGACTTTGGTAAGATTTTCAACATA 3891
    |||||
Db 214 CysLeuLeuValGlyCysAlaPheThrLysSerThrGlnPheGlyLeuHisIleTrpLeu 233
QY 3890 GGTGATGCTAGACTGCGCTCTAGCTAGCTAGTAAAGTCTGCTCTGTTCTTTCTTCCA 3831
    |||||
Db 234 ProAspAla---MetGluGlyProIleProValSer-----AlaLeuIleHisAla 249
QY 3830 AAGAAGTTCTACTG-----CTTGTTCTCATGAGTTTC-----TGTTCCTCTTTC 3786
    |||||
Db 250 AlaThrLeuValValCysGlyIleLeuLeuLeuSerPheValTyrTrpCysPheAspPhe 269
QY 3785 TCTTTTCA-----TATTGAATATACGTTTAAATGGTATATGTAATAATATCTCT 3732
    |||||
Db 270 TrpPheSerTyrPheTyrAsnLeuIleGlyTrpSerThrLeuIleLeuLeuMetThr 289
QY 3731 CTTTTTCTCTTTTAGGAGATGTTGCAATTCCTCTCAAGAAATGAATATCATTTGTTATC 3672
    |||||
Db 290 LeuCysValPhe-----TyrAsnPheAspValLysArgTyrValAlaPhe 304
QY 3671 TGTCTTTTGTGTCAGCTTTCTATGTCATGACTAATGCTGTGAAGCCACAAATTTTGTCT 3612
    |||||
Db 305 SerThrIleCysGlnIleSerPheSerMetPheCysCysLeuCysIleAspIleTyrIle 324
QY 3611 TGTAGTTGAATAATTTCTGACTAGAGACTCTGCTGTTTCAGGTGT----- 3567
    |||||
Db 325 GlySer-----LeuPhePheCysTyrHisMetPheTyr 335
QY 3566 -----TTGTTCCATTATCTTGTGCTTTTGATACATGTTCTGCTTCCCTTCATTTGAC 3516
    |||||
Db 336 LysAlaThrLeuPheIleValLeuGlyIleTrpIleHisIlePhePheGlyLeuGlnAsp 355
QY 3515 ACTGTTTTCAGTTGTCTCTTTGTGATGTTCTGAACCAATATTTCTCTTAGACATCT 3456
    |||||
Db 356 -Leu-----ArgCysTyrPhePheMetTyrPheCysGlyCysValLeuAlaArgLeuLeu 373
QY 3455 CTTGCTAATTGAGATAATTTTGTAGTTTGGATTTCTTCTTGAGCTCAGAAAGTGGTGATG 3396
    |||||
Db 373 uleuilePheAlaileLeu----- 379
QY 3395 GAGCACTCAATGTTAATATCGTACTCTCACATCAACATTCATTTCTTTGCAACAGCAT 3336
    |||||
Db 380 -----AsnSerCys 382
QY 3335 CTCTCAATGTGGAGCAGGTTCTTACTTTTCTTGATGTCACAATGATGATGCTGTTG 3276
    |||||
Db 382 sSer----- 383
QY 3275 ACAGCGAGCCAGTCAGGATGGTGATTCATTTCTGCTCTAGTTTCTTTGTTCTGTTT 3216
    |||||
Db 384 -----IleTrpPheLeuCys-- 388
QY 3215 TTCCTCAATTTAGATGACCATTTGTTCTCAGCTATCAGAACTTTAGTCGCCCATATT 3156
    |||||
Db 389 -----GlyPheTyrCys-----LysAspMetLeu 396
QY 3155 GAGATCCCTTTAGTAATGATTCCTTTTTCAGTTTGTAGGTCATCTAAGTTCAGCATCTTT 3096
    |||||
Db 396 uleuAlaLeuLeuMetLeuLeuSerPheTyr-----AsnIle-IleGluPheLeuP 413
QY 3095 TCTTTAAATCTTAAGTCTCAAGATTTATTTTCTCTT-----TTC 3054
    |||||
Db 413 heileSerIleile-----PheIlePhePheThrMetIleTyrAsnTyrPheLeu 429
QY 3053 TGGATTGTTGTTTAGTGTGGCTATTTCAG-----TTTGTGATGGCAATTCCTTTTCAACAT 2997
    |||||
Db 429 euLeuPhePheLeuMetPheValPheLysCysPheCysLeuVal-----AspC 445
QY 2996 GCAATTTCATGTAAGATAAATTTTTCATTTTCATGAGTTGAGAAACCTATTCAAAATTAATTTC 2937
    |||||
Db 445 ys-----LeuPheLeuLeuPheA 451
QY 2936 CTACACCTTCAATTTCTTATCTGTATCTGAGAGCCCTTCAAGTGTGTTTCACTTCTTAC 2877
```


| | | | |
|----|------|---|------|
| QY | 2351 | CTGCTTTGAAAGTTTCATGCCATCAATTCAGAGCTTTAGTGAATAGAACTTTTCATTC | 2299 |
| Db | 349 | yserIle----- | 351 |
| QY | 2291 | TCAGGAGGCGCTTCAGAAAAACCTCATTATCAGAGACTCTTCTATTTCCTCATTTATTTATCT | 2232 |
| Db | 352 | ----- | 356 |
| QY | 2231 | ATTCTTTTTCATGTGTAGCTGGGTAAACACCATCCTCTCTGGAACAGTCTACAGAGACT | 2172 |
| Db | 356 | yPhePheLeu----- | 359 |
| QY | 2171 | CCAGATCCCAAGAACTTCTCATAGTCCTTTGTTTGGATTTCATGGGAGTT----- | 2120 |
| Db | 360 | -----LysValPheCysLeuLeuHisLeuSerTyLeuGlyLeuCysIleVa | 376 |
| QY | 2119 | -----CTCATTCGCCTCAATGTGTTTCATTCCTTAATTCCAAGGTTTATTTGGAC | 2070 |
| Db | 376 | lPhePheSerIlellelPrLeullellelleTyrllePheTyrlPheArgLeullelleAs | 396 |
| QY | 2069 | AGACTTTTGTGATTCAAATGGAGGCTCGAAGCAGATGGCTCTCGGGAGGCCTGCTTTGAAG | 2010 |
| Db | 396 | nIlePheIlePheSerTyrlGlnPheIleGlyPheTrpValVal-----Ar | 411 |
| QY | 2009 | TTTGATGTGCATCAATTCCTAGGCTTTAGTTGGAAATAGAACTTTCATTCCCAGTTAGCCTC | 1950 |
| Db | 411 | gLeuHisIleLeuGlyPheSer-----AsnLeuPhePheIleLeu----- | 424 |
| QY | 1949 | AGAAGACCACTTAAACAGGGACCTCTAAATTTTCCATTATTTATCTATTTTGTAG | 1890 |
| Db | 425 | -----SerPheSerIlePheIleLeuPhePheAspIl | 435 |
| QY | 1889 | CGCAGCGCTTCGTTAAACACAATCCTT | 1865 |
| Db | 435 | elleAsnLeuPheAspLeulleLeu | 443 |

RESULT 9

D90129

hypothetical protein orf1613 [imported] - Guiliardia theta nucleomorph

C;Species: nucleomorph Guiliardia theta

A;Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont

C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001

C;Accession: D90129

R;Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu

Nature 410, 1091-1096, 2001

A;Title: The highly reduced genome of an enslaved algal nucleus.

A;Reference number: A99082; MUID:11323671; PMID:11323671

A;Accession: D90129

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1613 <DOU>

A;Cross-references: GB:AF083031; NID:g13794350; PIDN:AAK39727.1; GSPDB:GN00152

C;Genetics:

A;Gene: orf1613

A;Map position: 3

A;Genome: nucleomorph

C;Keywords: nucleomorph

| | | | |
|------------------------|----------|---------------|------|
| Alignment Scores: | | | |
| Pred. No.: | 0.000241 | Length: | 1613 |
| Score: | 136.00 | Matches: | 238 |
| Percent Similarity: | 34.45% | Conservative: | 195 |
| Best Local Similarity: | 18.93% | Mismatches: | 416 |
| Query Match: | 1.89% | Indels: | 409 |
| DB: | 2 | Gaps: | 61 |

US-09-602-362E-22 (1-4115) x D90129 (1-1613)

| | | | |
|----|------|--|------|
| QY | 4019 | GTGCAGCTTATGCAATTCGGACAAAGTAATCTTCACACTTATTATTATGCTAGGCTTC | 3960 |
| Db | 386 | LeuIlyPheLeuGlnPheCysAspLysAsnLeuHis-----Phe | 398 |
| QY | 3959 | AGGAAGAAAACATGAATTTCTCTTAAATAATAGTATTCGTGCACACAGACTTTGGTAAGATT | 3900 |

[illegible]

| | | | |
|----|------|--|------|
| Qy | 2295 | ATTCTCAGGAGCGCTTCAGAAAACCTCATATACAGAGACTCTTCTAT---- | 2249 |
| Db | 1079 | snPheLeuPheLeuPheHisLeuCysAspPheLysArgLeuPhePheLeillelleSerP | 1099 |
| Qy | 2248 | -----TTTCCATTATTTTATCTATCTATCTTTTGA | 2221 |
| Db | 1099 | heSerAsnLysValLysSerSerThrLeuPhePheThrTyrlleTyrPhe---- | 1117 |
| Qy | 2220 | TGTGTAGCCTGGGTAAACACCATCTCTCTGGAAACAGCTCACAGAGA---- | 2173 |
| Db | 1118 | -----AsnLysLeuValArgLysLeuValleP | 1127 |
| Qy | 2172 | TCCAGAATCCCAAGAACTTTCTC--ATAGTCCTCTGTTGGATTGATGGAGTCTCTCA | 2116 |
| Db | 1127 | heArgAsnLeuLysllePheilleArgLeuThrSerMetValLyslleMetGlyLeuLeuI | 1147 |
| Qy | 2115 | TCGCGCTCAATGTTTGTTCATCTTAATTCCAAGGTTTATTTGGACAGACTTTTGATTTC | 2056 |
| Db | 1147 | leArgPheAsnTyr-----PheArgPheLeu | 1155 |
| Qy | 2055 | AATGGAGGCTCGAAGGCAGATGGCTCTCGGAGGCGCTGTTTGAAGTTTGCATGTCA | 1996 |
| Db | 1155 | ysPheAsnLeu----- | 1158 |
| Qy | 1995 | ATTCTAGGCTTTAGTTGGAATAGAACTTTCATTTCCAGTTAGCCTCAGAAGACCACTT | 1936 |
| Db | 1159 | -----TyrgluilleGlyPheSerArgLysLysgluTyrgluLysAsnTyT | 1174 |
| Qy | 1935 | ACAGGACCCCTCTAATTTTCCATTATTTATCTATTTCTTTGAGCGCAGCTTCGTAA | 1876 |
| Db | 1174 | yrLysThrLysSerThrCysLysTyPhePheLysTyPheValLeuLysArglleleP | 1194 |
| Qy | 1875 | ACAAATC-----CTTCTGTGAACAGTCTCAGAGACTCCAGTATCCAGAAATTTCT | 1822 |
| Db | 1194 | heAspileAsnGlnlleLeuPheilleAlaThrAsnLysLeuLeuPheAsp----- | 1212 |
| Qy | 1821 | TATAGTCCTTTGTTGGATTCTCATGGGA-----GTTCTCATCTCGCTC---- | 1778 |
| Db | 1212 | alAsnArgPheilleGluilleLeuAsnAsnLeuAsnLysllelleThrTyrcysvallele | 1232 |
| Qy | 1777 | -----CATGTTTTTCATTTTTCATTTCCAGAGCTTAT | 1747 |
| Db | 1232 | ysLeuAsnLysLysThrGluLysAsnTyTyArgPheAsnPheValTyHisSerPheP | 1252 |
| Qy | 1746 | TTGGGACAGAC-----TTTTGATTTTCAGTGGAGGCTCGAAGCAGATGGCTTCC | 1699 |
| Db | 1252 | heGluArgAspLeuTyArgAsnPhePheLeuTyrlleuSerlleGlnMetLysSerPheP | 1272 |
| Qy | 1698 | CGGAGGCTCGCTTTGAAAGTTTGCATGCTCTCAATTCTAGGCTTTAGTGAATAGAAAT | 1639 |
| Db | 1272 | heCysSerHisLeuSerSerLysAsnilleAsnAsnSerThrVal-----ThrLysV | 1289 |
| Qy | 1638 | TCATTCCGAGTAGCCTTCAGAGACACATTTATTAGG----- | 1601 |
| Db | 1289 | alCysLysLysLysLysSerLysSerHisPheilleArglleLysLeuAsnLyslleTyA | 1309 |
| Qy | 1600 | -----GACTCTCTATTTTCCATTTTTTATCTATTTTTCATGTAGCCTTGG | 1549 |
| Db | 1309 | spArgPheAspLysThrilleLeuHisThrilleleTySer-----V | 1323 |
| Qy | 1548 | TAAACACATCTCTGTAAACAGCTCTCAGA-----GACTCTAGAA | 1507 |
| Db | 1323 | alThrAsnTyPheilleLysLysCysAspLysTyMetLeuMetPheTyrlLysPheLeuM | 1343 |
| Qy | 1506 | TCCACAGAATTTCTTATAGTCTCTTTGTTGG-----ATTCCGTGGGACACGGAT | 1456 |
| Db | 1343 | etLysArgillePheLeuAspSerValAsnTyrlLyslleSerTyrgluillePheP | 1363 |
| Qy | 1455 | CTGCTTCAATGTTTGTTCATCTTCATTCAAAGGCTTATTTGGAAGAGTTTTCCTTCA | 1396 |
| Db | 1363 | hePheGlnlleSerValVallleSerPheSerAsnLeuPheProlLysTytr-----A | 1380 |

QY 2544 GCGTGGGTACACACCATCTCTGTGAAACAGTCCACGGAGACTTCAGATCCCGAGATT 2485
Db 434 hePheGly-----ProLeuVal-----P 440
QY 2484 TTCTCAACCTCTTTGTTTGTATCTCTGAAGGAAATCTGATCTGTCCTCAATGTTGTTTCAT 2425
Db 440 heLeuPheMetMetilePhe-----LeuSerP 449
QY 2424 TCTTAATTCGAAGGTTATTTGGAACACAGACTTTTCATTTCAATGAGCTCGAAGCAGAT 2365
Db 449 heLeuileLeuLysMetLeuPheLysGluLeuMetTyrLysPheLeuVal-----AspT 467
QY 2364 GCGCTCTGGGAGGCTGCTTTGAAGTTTGCATGCCATCAATTCAGAGCTTTAGTGGAT 2305
Db 467 yrLeuAlaLysAsnSerileTyrLysMetLysAsnLeuLysPheMetAspLeuPheLeuA 487
QY 2304 AGAATTTTCATCTCAGGGA-----GCCTTCAGAAAACCTCATTT 2266
Db 487 snAsnileAsnSerLysGlyTyrThrLeuPheLeuSerSerGlyMetPheLysAsnTyrT 507
QY 2265 ATCAGAGACTCTCTATTTT-----CCATTTATTTATCTATTTCTTTTGTATG 2218
Db 507 yrLeuLysSerLeuAsnPheAsnSerValValLeuilePheilePheMetileC 527
QY 2217 GT 2216
Db 527 ys 527
RESULT 12
T44522
hypothetical protein 4S [imported] - Shigella sonnei
C:Species: Shigella sonnei
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T44522
R:Chida T.; Okamura, N.; Yoshida, Y.; Ohtani, K.; Arakawa, E.; Watanabe, H.
submitted to the EMBL Data Library, May 1999
A:Description: Complete DNA sequence of the O-antigen (rfb) gene cluster in Plesiomonas
A:Reference number: 222787
A:Accession: T44522
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-426 <CHI>
A:Cross-references: EMBL:AB028135; PIN:BA05073.1
A:Experimental source: isolate HW383
Alignment Scores:
Pred. No.: 0.000367 Length: 426
Score: 133.00 Matches: 108
Percent Similarity: 37.42% Conservative: 78
Best Local Similarity: 21.73% Mismatches: 169
Query Match: 1.85% Indels: 143
DB: 2 Gaps: 21
US-09-602-362E-22 (1-4115) x T44522 (1-426)
QY 3856 GTAAAGTCTGGTCTGTTGTTTCTCCAAAGAGTTTCTACGTCTG-----TCGGCA 3812
Db 6 ileAspAlaGlyGlyThrPheLeuLeuLysAlailePheGlnileGlyValPheValTyr 25
QY 3811 TTCTCATGAGTTCT-----GTTCTCTCTCTCTCTTTTTCATATTTGAATATACGTTTTT 3758
Db 26 PheThrHisValSerAspilleThrPheGlyileileSerTyrValPheThrValTyr 45
QY 3757 AATGGATTGTAATTAATATCTCTCTTTTCTTTTATAGGAGATGTGTGCAATTCCT 3698
Db 46 ---TrpPhe-ValLeuAsnPheSerAspTyrGlyPheArgThrLysLeuValLysAsp11 64
QY 3697 CTCAGAAAGTAATATCATTTGTTATCTGCTTTGTTGTCAGCTTCTATGTGTCATGACTA 3638
Db 64 eSerAspAsnSerTyrSerAlaSerGluLeu----- 74
QY 3637 ATTGCTGTGAAGCACCAAAATTTTGTGTAGTTGAAATAATTTCTGCTAGACTCTGC 3578

75 -----LeuSerArgSerAspG1 80
3577 TGTTCAGGTCTTTGTTCCATTATCTTGTCTGTTTGTATACATGTTGCTTCTCTTCATG 3518
Db 80 yValLysThrTyrVal-PhePhePheilePheilePheMetPheTyrSerTyrVal 100
QY 3517 ACATGTTTTCAGGTCTCTCTTGTGATGTTCTCGAAACCAA-----TGTATT 3470
Db 100 erAspSerileSerLeuThrLeu-LeuValTyrileSerSerAlaTyrPheValCysile 119
QY 3469 TCTCTTAGACATCTCTGCTTAATTTAGATAAATTTTGTGATTTTGGATTTCCTTTGAGCTC 3410
Db 120 SerSerGlyArgPheSerLeuLeuGlnAlaValGlyArgPhe-----ArgCysGluLeu 137
QY 3409 AGAAAGTGGTATGGAGCACTCATTTTAATATCGTACTCTCACATCAACATTCATTTTC 3350
Db 138 -----TyrileAsnileTyrSerThrileileTyr 147
QY 3349 TTTGCAACAGCATCTCCTCAATGTGGAAGCAGGTCTTCTTTTCTTGATGTCACAT 3290
Db 148 -ile-----GlyCysAsn-----LeuPheLeuSerLeuPheile 158
QY 3289 TGATCATGGCTGTACAGCGAAGCCAGTCAGATGGTGTATTCATTTCAATTTCTGCC----- 3237
Db 158 eGluProLeuTyrTyrSerAlaileSer-IlePheileTyrSerileSerLeuLeuValP 178
QY 3236 --TCTAGTTTCTTGTCTGTTTCTTCTCAATTTAGATGAGCATTTCTCAGCATTC 3179
Db 178 heSerSerHisLysCys----- 183
QY 3178 AGAACTTTAGCTGCCACATATTTAGATCCCT-----TTTA 3143
Db 184 --AsnValProCysPheHisileLysArgProSerileLeuValTyrLysAspPheLeuA 203
QY 3142 GTAATGATTCCTTTTTCAGTTTGGGTCAATCAAGTTCAGCATCTTTCTTTTAAATCTT 3083
Db 203 spAlaThrProPheAlaileLeuValLeuLeuAsnValValLeuSerSerileAspLeuP 223
QY 3082 AAGTCTCTCAAGATTATTTTCTCTTCTGATGGTGTGTTAGTGTGCTATTCAGTT 3023
Db 223 heileLeuLysGluTyrPheSer-----TyrAsnSerV 234
QY 3022 TTGCATGGCAATTCCTTTTCAACATGCAATTTCAAGAAATAATTTTCATTCATCA 2963
Db 234 alAla-----ileTyrGlnValValThrArgValAsnThrGlyLeuileileValP 251
QY 2962 GTTGAAACCTATTCAAAATTTACTTTTACACTTCAATCTTATCTGTATCTGAGAGC 2903
Db 251 heAsnValileTyrThr-ValLeuLeuProSerPheSerTyrTyrLeuLysAsnSerGlu 270
QY 2902 CTTTCAAGTGTGTTTCTCACTTCTTCTATGCTGCTCTCGATTTCTTCTAATTTCTT 2843
Db 271 TrpGlyAsnileArgLysLeuGlnArgTyrileSerLeuLeuValLeuLeuCysLeu 290
QY 2842 CCTAATTTTCTTTTAAATATA-----TCGGCA 2816
Db 291 CysTyrTyrPhePheGlyileTyrPheValGlyileLeuPheGlyAspGluTyrLysVal 310
QY 2815 TTTCTCTCTCTCTTCTTCTTCTGTTTAAAGTCAATCTACACTGCAGACTCTGTTCCCATTT 2756
Db 311 ileSerSerAlaThrPheLeuile-----MetPheMetAlaLeuileLysTyr 326
QY 2755 AACTTTTGGTCTCTACTGTGATTTTCTTTTCTGACAGCTTCTTTTTCAGTACAC 2696
Db 327 AsnPheTyrPheLeuileAsnGluLeuTyrLeuValCysSerGlyAsnGlnSerGluArgVal 346
QY 2695 AAAACT-----CTTTTCTCATTTTGTCCATTTTCTGTCAGTGTTCACA---- 2654
Db 347 LysSerTyrCysileGlyValileSerMetAlaValPhePheTyr-PheileProArg 366
QY 2653 -----GTGCTTTTGAAGTTTCCCTTGC-----TTT 2627

[illegible]

Search completed: July 15, 2004, 09:48:56
Job time : 119.145 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 08:50:37 ; Search time 192.589 Seconds

(without alignments)
13356.780 Million cell updates/sec

Title: US-09-602-362E-22

Perfect score: 7086

Sequence: 1 ctactctatcacgaacagc.....gagtggaactccactggaaa 4115

Scoring table: BLOSUM62

| | |
|-------------|-------------|
| Xgapop 10.0 | Xgapext 0.5 |
| Ygapop 10.0 | Ygapext 0.5 |
| Fgapop 6.0 | Fgapext 7.0 |
| Delop 6.0 | Delext 7.0 |

Searched: 1285345 seqs, 312560633 residues

Total number of hits satisfying chosen parameters: 2570690

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US09602362/runat_15072004_093627_22071/app_query.fasta_1.10325
-DB=Published Applications AA -QFMT=fastan -SURFIX=rapb -MINWATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-NCPU=200000000 -USER=US09602362 -CGEN_1_1 7.9 @runat_15072004_093627_22071
-LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database : Published Applications AA:*

| | |
|-----|--|
| 1: | /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.* |
| 2: | /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.* |
| 3: | /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.* |
| 4: | /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.* |
| 5: | /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.* |
| 6: | /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.* |
| 7: | /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.* |
| 8: | /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.* |
| 9: | /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.* |
| 10: | /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.* |
| 11: | /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.* |
| 12: | /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.* |
| 13: | /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.* |
| 14: | /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.* |
| 15: | /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.* |
| 16: | /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.* |
| 17: | /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.* |
| 18: | /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
|------------|-------|-------------|--------|-------|-------------|

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 |
|-------|------|------|----|---------------------|-------------------|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1234 | 17.4 | 1341 | 12 | US-10-058-270A-4 | Sequence 4, Appli | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1234 | 17.4 | 1341 | 13 | US-10-007-805-565 | Sequence 565, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1234 | 17.4 | 1341 | 14 | US-10-076-622-565 | Sequence 565, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1234 | 17.4 | 1341 | 15 | US-10-177-293-334 | Sequence 334, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1234 | 17.4 | 1341 | 16 | US-10-124-805-565 | Sequence 565, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1223 | 17.3 | 1349 | 13 | US-10-007-805-573 | Sequence 573, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1223 | 17.3 | 1349 | 14 | US-10-076-622-573 | Sequence 573, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1223 | 17.3 | 1349 | 15 | US-10-124-805-573 | Sequence 573, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 926.5 | 13.1 | 1013 | 13 | US-10-007-805-553 | Sequence 553, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 926.5 | 13.1 | 1013 | 14 | US-10-076-622-553 | Sequence 553, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 926.5 | 13.1 | 1013 | 15 | US-10-124-805-553 | Sequence 553, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 917 | 12.9 | 1002 | 9 | US-09-604-287A-475 | Sequence 475, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 917 | 12.9 | 1002 | 10 | US-09-551-621-475 | Sequence 475, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 917 | 12.9 | 1002 | 11 | US-10-007-805-475 | Sequence 475, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 917 | 12.9 | 1002 | 12 | US-10-076-622-475 | Sequence 475, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 914.5 | 12.9 | 1095 | 13 | US-10-124-805-493 | Sequence 493, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 914.5 | 12.9 | 1095 | 14 | US-10-007-805-493 | Sequence 493, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 914.5 | 12.9 | 1095 | 15 | US-10-076-622-493 | Sequence 493, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 612.5 | 8.6 | 1239 | 13 | US-10-124-805-433 | Sequence 433, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 612.5 | 8.6 | 1239 | 14 | US-10-007-805-577 | Sequence 577, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 612.5 | 8.6 | 1239 | 15 | US-10-076-622-577 | Sequence 577, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 597 | 8.4 | 661 | 13 | US-10-124-805-577 | Sequence 577, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 597 | 8.4 | 661 | 14 | US-10-007-805-552 | Sequence 552, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 597 | 8.4 | 661 | 15 | US-10-076-622-552 | Sequence 552, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 583.5 | 8.2 | 650 | 9 | US-10-124-805-552 | Sequence 552, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 583.5 | 8.2 | 650 | 10 | US-09-825-301-25 | Sequence 25, Appl | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 583.5 | 8.2 | 650 | 11 | US-09-604-287A-469 | Sequence 469, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 583.5 | 8.2 | 650 | 12 | US-09-551-621-469 | Sequence 469, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 583.5 | 8.2 | 650 | 13 | US-10-007-805-469 | Sequence 469, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 583.5 | 8.2 | 650 | 14 | US-10-076-622-469 | Sequence 469, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 583.5 | 8.2 | 650 | 15 | US-10-124-805-469 | Sequence 469, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 583.5 | 8.2 | 650 | 16 | US-10-033-527-25 | Sequence 25, Appl | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 582 | 8.2 | 743 | 13 | US-10-007-805-494 | Sequence 494, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 582 | 8.2 | 743 | 14 | US-10-076-622-494 | Sequence 494, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 544.5 | 7.7 | 1225 | 14 | US-10-124-805-494 | Sequence 494, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 481 | 6.8 | 1011 | 16 | US-10-177-293-332 | Sequence 332, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 465 | 6.6 | 512 | 16 | US-10-408-765A-1557 | Sequence 1557, Ap | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 421.5 | 5.9 | 445 | 9 | US-10-181-663-16 | Sequence 16, Appl | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 421.5 | 5.9 | 445 | 10 | US-09-825-301-29 | Sequence 29, Appl | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 421.5 | 5.9 | 445 | 11 | US-09-604-287A-473 | Sequence 473, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 421.5 | 5.9 | 445 | 12 | US-09-551-621-473 | Sequence 473, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 421.5 | 5.9 | 445 | 13 | US-10-007-805-473 | Sequence 473, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 421.5 | 5.9 | 445 | 14 | US-10-076-622-473 | Sequence 473, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 421.5 | 5.9 | 445 | 15 | US-10-124-805-473 | Sequence 473, App | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 421.5 | 5.9 | 445 | 16 | US-10-033-527-29 | Sequence 29, Appl | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

ALIGNMENTS

RESULT 1

US-10-058-270A-4

; Sequence 4, Application US/10058270A

; Publication No. US20040029114A1

; GENERAL INFORMATION:

; APPLICANT: Mack, David H.

; APPLICANT: Gish, Kurt C.

; APPLICANT: Afari, Daniel

; APPLICANT: Eos Biotechnology, Inc.

; TITLE OF INVENTION: Methods of Diagnosis of Breast Cancer, Compositions and Methods of Screening for Modulators of Breast Cancer

; FILE REFERENCE: 018501-005210US

; CURRENT APPLICATION NUMBER: US/10/058,270A

; CURRENT FILING DATE: 2002-01-24

; PRIOR APPLICATION NUMBER: US 60/263,965

; PRIOR FILING DATE: 2001-01-24

; PRIOR APPLICATION NUMBER: US 60/265,928

; PRIOR FILING DATE: 2001-02-02

; PRIOR APPLICATION NUMBER: US 09/829,472

; PRIOR FILING DATE: 2001-04-09

; PRIOR APPLICATION NUMBER: US 60/282,698

; PRIOR FILING DATE: 2001-04-09

; PRIOR APPLICATION NUMBER: US 60/288,590

; PRIOR FILING DATE: 2001-05-04

```

; PRIOR APPLICATION NUMBER: US 60/294,443
; PRIOR FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1341
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-058-270A-4

Alignment Scores:
Pred. No.:      8,89e-110      Length:      1341
Score:          1234.00        Matches:     679
Percent Similarity: 59.18%      Conservative: 146
Best Local Similarity: 48.71%    Mismatches:  376
Query Match:      17.41%       Indels:      207
DB:               12          Gaps:         115

US-09-602-362E-22 (1-4115) x US-10-058-270A-4 (1-1341)
QY 96 CAAGGGAAGAGACATCAACCTTAATATACAGAG---CCAGAGAGAGCTGCTCTAAC 152
DB 3 LysArgLysLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu-- 21
QY 153 TGGGCTGCTCAATGCGCT-----GAGGAGTACTAAATTC---GGTA 194
DB 22 -----HisTrpAlaCysValAsnGlyHisGluGluValValThr-PheLeuValA 38
QY 195 GACAGAGTCCAGCTGACCTCTTC---TGGCGAACACGACACCTCTGTATGAGGC-T 250
DB 38 spArgLysCysGlnLeuAspValLeuAspValLeuHisArgThrProLeuMetLysAla 58
QY 251 TACAATGCCACAGAGGCTTTC---AAATATTGATAGATTC-CGTCCGATATAATC 306
DB 58 euGlnCysHisGlnGluAlaCysAlaAsnIleAspSerGlyAlaAspIleAsnL 78
QY 307 TCGTG---ATGTGTATGCAACATGGCTTCCATTATGCGTTTATAGTGGATTGTCAGT- 362
DB 78 euValAspValTyGlyAsnMetAlaLeuHisTyAlaValTySerGluIleLeuSerV 98
QY 363 --GGTGGCAACTGCTGTC-CATGGTGCAGTATCGAAGTC---CAACAGGCTGCTCA 416
DB 98 alValAlaLysLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeu 118
QY 417 CACCACCTTTACTACCAT---AACCAAGAAGTGAAGCAATTGTGGAATTTT---GCTGAT 470
DB 118 hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuLeu 137
QY 471 AAAATGCAATCGAATCGTTA---ATAAGTTAAATGCCACACCTCATGCTGCTGAT 527
DB 138 LysAsnAlaAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAla 157
QY 528 G---TCTGGATCATCAGATAGTTGTCATGCTTCTTAGCAAAA---TGTGACGCTTT-- 579
DB 157 lCysHisGlySerSerGluIleValGlyMetLeuLeuGlnAsnValAspValPheAl 177
QY 580 -GTGAGATATAGTGAATCACTGACAGACAT---ATGCTGTACTGTGGATTCTCACA 635
DB 177 alAlaAspIle-CysGlyValThrAlaGluHisTyAlaValThrCysGlyPheHisI 197
QY 636 TTCATACAAA---TTATGAATATATACGAAATATCAAAATATCA---TCAAAATACAA 689
DB 197 leHisGluGlnIleMet-GlutyrIleArgLysLeuSer-LysAsnHisGlnAsnThrAs 216
QY 690 T---CAGAGGAACCTGCGAGNAACCTGATGAGCTGCACCT---GGCGGAAGACAC 743
DB 216 nProGluGlyThrSerAlaGlyThrProAspGluAlaAla-ProLeuAlaGluArgThr 236
QY 744 CTGACACGCTGAAAGCT---GGTGGAAAACACCTGATGAGCTGCACCT---GGTGA 797
DB 236 roAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAla-ProLeuValGlu 255
QY 798 AAGACACCTGACAGGCTGAAAGCT---GGTGGAAAACACCTGATGAGCTGCATCT--- 851

```

```

DB 256 ArgThrProAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAlaSerLeu 275
QY 852 TGGTGGAGGAACATCTGACAAATCAATGTTGGAGAAAGG---ACATCTGGAAGTTCGA 908
DB 276 Val-GluGlyThrSerAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheGl 295
QY 909 AGCTCAGC---AGAAAACACCTAGGAATACGATCC---TGCACAAAACATCTCAGA 962
DB 295 uGln-SerAlaGluGluThrProArgGluIleThrSerProAlaLysGluThrSerGlu 315
QY 963 AATTACG---GCCAGCAAGAGACCTAGAGATCGCAGGAGAA---AAAGAAGAC 1016
DB 315 ysPheThrTrpPro-AlaLysGlyArgProArgLysIleAlaTrpGluLysLysGluAsp 334
QY 1017 ACACCTAGGGAATATGATGCCGAAAAGAAACACTGAGAAA-----TTA 1061
DB 335 ThrProArgGluIleWetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAla 354
QY 1062 CBTGGCAGAAAAGAGAACCTTAGGAAGCCATGGGAGAAAAGAAA---CCTGTGAAA 1118
DB 355 LysGlyArgProArgLysIle-----AlaTrpGluLysLysGluThrProValLys 371
QY 1119 GATGATGCT-GCAAGAGTACATCTAATAACTAAAG---TTTGAAGAAAGAAATCT 1174
DB 372 ThrGlyCysValAlaArgValThrSerAsnLysThrLysValLeuGluLysGlyArgSer 391
QY 1175 AAGATGTT-GCATGTCCACAAAAG---AATCATCTACAAAGCAAGTCCATGATCAGAG 1230
DB 392 LysMetIleAlaCysProThrLysGluSerThrLysAlaSerAlaAsnAspGlnArg 411
QY 1231 ---TCCCATCATCCAAACAGGGAAGATCAAGATATCTTT-GATTCTCGGCTCT 1286
DB 412 PheProSerGluSerLysGlnGluAspGluGlySerCysAspSerArg---Ser 430
QY 1287 TTG-----AGTTCTGCAAGATTCAAGTG---GTATACCTGAGTCTATATATAAAGT 1337
DB 431 LeuPheGluSerSerAlaLysIleGlnValCysIleProGluSerIleTyGlnLysVal 450
QY 1338 AATGAGATAATAAAGAG---TAGAAGCCCTCTAAAGCCATCTGCTTCAAGCCGCCA 1394
DB 451 MetGluIleAsnArgGluValGluGluProLysLysProSerAlaPheLysProAla 470
QY 1395 TTGAAGCAAAA---CTCTTTCCAAATAAGCCTTTG-----AATGAAGAAATGAACA 1442
DB 471 IleGluMetGlnAsnSerValProAsnLysAlaPheGluLeuLysAsnGluGln----- 488
QY 1443 AACATTGAAGCAGATCCGTG-TTCCACCG---AATCCAAACAAAGGACTATAAGAAAAT 1498
DB 489 ThrLeuArgAlaAspProMetPheProProGluSerLysGlnLysAspTyrGluGluAsn 508
QY 1499 TCTTGGGATTCAGAGTCTCTG---GAGACTGTTTACAGAGG---ATTGTGTTACCAA 1552
DB 509 SerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLy 528
QY 1553 GCCTACACATCAAAAAGAAATAGATAAAA---AATGGAAAA---TAGAAGAGTCCCTAAT 1606
DB 528 salThrHisGlnLysGlu-IleAspLysIleAsnGlyLysLeuGluGluSerProAsnL 548
QY 1607 AAAGTGTCTTCTGAGGCTACCTCGGAAT---GAAATTTCTATTCACCTAAAGCCT--- 1660
DB 548 ysAspGlyLeuLeuLysAlaThrCys-GlyMetLysValSerIleProThrLysAlaLeu 567
QY 1661 ACAATTGAAGACATCAAACTTCAAGAGGAGCCTCC---GGGAAGCCTCTGCT---TC 1714
DB 568 GluLeuLysAspMetGlnThrPheLysAlaGlu-ProProGlyLysProSerAlaPheGl 587
QY 1715 GAGCCTCCACTGAATCAAAAGTCTGTCCCAATTAAGCCTTGG-----AATGAAA 1765
DB 587 uProAlaThrGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu-- 606
QY 1766 ATGAAAAACATGGAGGAGATGAGAA---CTCCCATCAGAAATCCAAACAAAGACTATAA 1822

```

Db 607 ---GlnThrTrpArgAlaAsp-GluLeuLeuProSerGluSerLysGlnLysAspTyrG 625
QY 1823 GAAAA---TTCTGGGATCTGGAGTCTCTG---TAGACTGTTTTCACAGAAGGATTG-TGT 1875
Db 625 LuGluAenSerTrpAspThr-GluSerLeuCysGluThrValSerGlnLysAspValCys 644
QY 1876 TTACCAAGG---CTGCGCTCAAAAGATAGATAAATAATGGAATAATAGAGGTT-- 1929
Db 645 LeuProLysAlaAlaHisGlnLysGluLeuLeuAspLysLeuGluGlySer 664
QY 1930 CCCTGTTAAAGTGTCTTCTGAGCTAACT---GGGAATGAAAGTCTTATTCCAACTAAA 1986
Db 665 ProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLys 684
QY 1987 GCCTAGA---ATTGATGACATGCAAACTTCAAAG---CAGCGCTCCGACAGCATCTGC 2040
Db 685 Ala-LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAl 704
QY 2041 CTTCGAGCTCCATTGAAA---TCAAAAGTCTGCCAAATAAAG---CTTGGAAATTAAGA 2094
Db 704 aPheGluProAlaIleGluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLysA 724
QY 2095 ATGAACAAACATTGAGGAGATGAGAA---CTCCCATCA---AATCCAAACAAAGGACTA 2148
Db 724 snGluGlnThrLeuArgAlaAsp-GluIleLeuProSerGluSerLysGlnLysAspTyr 743
QY 2149 TGAGAAAGTCTTGGGATCTGGAGTCTCTG---TAGACTGTTTCCAGAAGGATGGTG-- 2203
Db 744 GluGluSerSerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCy 763
QY 2204 -TTTACCAGCTACATCAAAAGAAATAGATAAAA---TAATGGAAATAGAGAGTC 2259
Db 763 sLeuProLysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluLys 783
QY 2260 TCT-GATAATCAGGTTTCTCAAGGCTCCCT---GAGAAATGAAAGTCTTATTCACATAA 2315
Db 783 rProAspAsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLy 803
QY 2316 GCCT---TGAATTGATGGCATCAAACTTTCAAAGCAGGCTCCCG---AGAGCCATCTG 2369
Db 803 sAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAl 823
QY 2370 CTTCGAGCCTGCATTGAAAT---GAAAAGTCTGTTCCAAATAAAC---CTTGGAAATTAAG 2423
Db 823 aPheGluProAlaIle-GluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLys 842
QY 2424 AATCAACAACATTGAGACATCAGAT---TTCCCTTCAGATCAAAACAAAGAGG 2477
Db 843 AsnGluGlnThr---LeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLys 861
QY 2478 TTGAGAAA---TTCTGGGATCTGAAGTCT---CCGTGAGCTGTTTCACAGAAGGATGG 2531
Db 862 ValGluGluAenSerTrpAspSerGlu-SerLeuArgGluThrValSerGlnLysAsp-V 881
QY 2532 TGTCTACC---CAGCTACACAAAAGAAAGATGAAAT---AGTGGAAATTAAGAAG 2585
Db 881 alCysValProLysAlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluA 901
QY 2586 ATTCACTAGCCTATCAAAAT---CTTGGTACAGTTTCAT---CTTGTGAAAGCGAAGGAA 2639
Db 901 spSerThrSerLeuSerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluL 921
QY 2640 CTTCAAAAGACATCTGGAACA-CGTACAGAAATGGAACAAT---GAAAAGAGATTIT 2695
Db 921 euGlnLysAspPheCysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheC 941
QY 2696 GTGTACTGAAAAAGAAC---TGTCAAGAACAAAGAAATAATAATCACAGTAGAGAA---CCA 2749
Db 941 ysValLeuLysLysLysLeuSerGluAlaLysGluIleLysSerGln-LeuGluAenGln 960
QY 2750 AAAGTTAAATGGGAACAGAG---TCTGCAGTGTAGATTGACTTTAAACCAA---GAAAAG 2803
Db 961 LysValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAenGlnGluGlu 980

QY 2804 AGAAGAGAGAAATGCCGATATATTAAAGAAAAA-----TTAGGAAGAAATTAG 2851
Db 981 LysArg-ArgAenAlaAspIleLeuAenGlnLysIleArgGluGluLeuGlyArgIle-G 1000
QY 2852 AAGAATCGAGACAGCAGATAGAAAGAGTGTAGAAGTGAACACACAACTTTGAAG---GCTCTC 2908
Db 1000 LuGluGlnHisArg-----LysGluLeuGluValLysGlnGlnLeu-GluGlnAlaLeu 1017
QY 2909 AGATCAAG---ATAGAATGAAGGTGTAGAAAGTAATTTCAATAGGTTTCTCA---AC 2962
Db 1018 ArgGlnAspIleGluLeuLysSerValGluSerAsnLeuAenGln-ValSerHisTh 1037
QY 2963 TCATGAATGAAA---ATTATCTTACATGAATTCATGTTGTAAGAAAGGATTCGCA-- 3017
Db 1037 rHisGluAenGluAenTyrLeuLeuHisGluAenCysMetLeuLysLysGluIleAlaLe 1057
QY 3018 -TGCAAAAGCTGAATAGCCACACTA---AACACCAATCCAGGAAAGGAAAAATAATC-T 3072
Db 1057 tLeuLysLeuGluIleAlaThrLeuLysHis-GlnTyr-GlnGluLysGluAenLysTyrP 1077
QY 3073 TTGAGGACTTAGATTTTAAAGA---AAAGATGCTGAACCTTAG-ATGACCTTAAACTGA 3128
Db 1077 heGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuMetThrLeuLysLeuL 1097
QY 3129 AAAGGAATCATTTACTAAAG---GGATCTCAATATGTGGCGAGCT---AAAGTTCTGATA 3182
Db 1097 ysGluGluSerLeuThrLysArgAlaSerGlnTyrSerGly-GlnLeuLysValLeuIle 1116
QY 3183 GCTGAGAAACAATGCTCATTTCTTAATTTGAGGAAACACAGAAAGAAA-----CTAGA 3236
Db 1117 AlaGlu-AsnThrMetLeuThrSerLysLeuLysGluLysGlnAspLysGluIleLeuG 1136
QY 3237 GGCAGAAATTGAATA-CACCATCCTGACTGGCTTCGCTGTACAAGC-----CA 3283
Db 1136 uAlaGluIleGluSerHisHisPro-----ArgLeuAlaSerAlaValGlnAspHi 1153
QY 3284 TGATCAATTTGACATCAAGAAAAAGTAAGA---ACCTGCTTCCATTGAG---GAGAT 3337
Db 1153 sAspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl 1173
QY 3338 GCTGTTTCAAGAAATGAATGTTGATGTG---AGAGTACGATATTACATAGTGTCT 3394
Db 1173 aCysLeuGlnArgLysMetAsnValAspValSerThrIleTyr-AsnAenGluValL 1193
QY 3395 CCATCACCACCTTTC---TGAGCTCAAGGAAATCCAAACCTCAAAATTTATCT---CAAT 3448
Db 1193 euHisGlnProLeuSerGluAlaGlnArgLysSerLysSerLeuLysIleAsnLeuAenT 1213
QY 3449 TAGCAGGAGATGCTTAAGAGA---AATCATTTGGTTTTCAGAACATCACAAAGAGA---CA 3502
Db 1213 yrAlaGlyAsp-AlaLeuArgGluAenThrLeuValSerGluHis-AlaGlnArgAspGl 1232
QY 3503 AGCTGAAAAACAGTGTCAAT---GAAGGAAGCGAACACATGTATCAAAAC-----GACAA 3553
Db 1232 nArgGluThrGlnCysGlnMetLysGluAlaGluHisMetTyrGlnAenGluGlnAspAs 1252
QY 3554 GATAATGGAAACAAACACCTGAACACAGCAGAGTCTCTAGTCAGAAATATTTCACATCAAG 3613
Db 1252 nValAsnLysHisThrGluGlnGlnSerLeu-AspGlnLysLeuPheGlnLeuGlns 1272
QY 3614 CAAAATTTGGCTTCACAGCAATTT---AGTCATGCACATAGAAAGCTGACACAAAG 3669
Db 1272 erLysAenMetTrpLeuGlnGlnLeuValHisAlaHisLys-----LysA 1288
QY 3670 CAGATAACAAT-----GATATTTCATTCTTGT---AGAGGAAATGCAACAC 3711
Db 1288 laAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLysMetGlnHisH 1308
QY 3712 ATCTCTTAAACAGAAAAAGAGGAGATATT---AATTACAATACCATTTAAAAACCGTAT 3768
Db 1308 isLeuLeuLysGluLysAenGlu-GluIlePheAsnTyrAsnAsnHisLeuLysAsnArg 1327

```

QY 3769 ATT---CAATATGAAAAAGACAAA 3789
Db 1328 lIeYrGlnTyrGluLysGluLys 1335

RESULT 2
US-10-007-805-565
; Sequence 565, Application US/10007805
; Publication No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007.805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 565
; LENGTH: 1341
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-007-805-565

Alignment Scores:
Pred. No.: 8,89e-110 Length: 1341
Score: 1234.00 Matches: 679
Percent Similarity: 59.18% Conservative: 146
Best Local Similarity: 48.71% Mismatches: 376
Query Match: 17.41% Indels: 207
DB: 13 Gaps: 115

US-09-602-362e-22 (1-4115) x US-10-007-805-565 (1-1341)
QY 96 CAAGGGAGAGACATCAACCTTAATATACAGAG---CCGAGAGAGACTGCTCTAAC 152
Db 3 LysArgLysLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu-- 21
QY 153 TGGGCTGGTCAATGGCCT-----GAGGAAGTAGTAACATTTTC---GGTA 194
Db 22 -----HisTrpAlaCysValAsnGlyHisGluGluValValThr-PheLeuVala 38
QY 195 GACAGAACTGGCAGTGCCTTGTG---TGGCGAACCAGACACCTCTGATGAAGGC-T 250
Db 38 sPArgLysCysGlnLeuAspValLeuAspGlyGluHisArgThrProLeuMetLysAla 58
QY 251 TACAAATGACAGAGAGGCTTTGTC---AAATATTTGATAGATTC-GGTCCGATATAATC 306
Db 58 euGlnCysHisGlnGluAlaCysAlaAsnIleLeuIleAspSerGlyAlaAspIleAsnL 78
QY 307 TCGTG---ATGTGTATGCAACATGGCTTCCATTATGGGTTTATAGTGGATTGTCAGT- 362
Db 78 euValAspValTyrGlyAsnMetAlaLeuHisTyrAlaValTyrSerGluIleLeuSerV 98
QY 363 ---GGTGGCAAACTGGTGC-CATGTGCAGTATCGAAGTGC---CAACAAGGCTGCCTCA 416
Db 98 alValAlaLysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeu 118
QY 417 CACCACTTTTACTACCAT---AACGAAAGAGTGAAGATTTGTGGAATTTT---GCTGAT 470
Db 118 hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuLeu 137
QY 471 AAAATGCAAAATGCGAATGCGTTA---ATAAGTTAAATGCACACCTCATGCTGCTAT 527
Db 489 ThrLeuArgAlaAspProMetPheProGluSerLysGlnLysAspTyrGluGluAsn 508

138 LysAsnAlaAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAlaVa 157
528 G---TCTGGATCATCAGAGATAGTTGTCATGCTTTAGCAAAA---TGTGACGCTCTTT-- 579
157 lCysHisGlySerSerGluIleValGlyMetLeuLeuGlnGlnAsnValAspValPheAl 177
580 -GTGCAGATATAGTGGAGTAATCTGCAGAACAT---ATGCTGTTACTGTGGATTTCTCACA 635
177 aAlaAspIle-CysGlyValThrAlaGluHisTyrAlaValThrCysGlyPheHisIle 197
636 TTTCAATAACAAA---TTATGAATATATAGAAAATTTATCAAAATCA---TCAAAATACAA 689
197 lHisGlnIleMet-GluTyrIleArgLysLeuSer-LysAsnHisGlnAsnThrAs 216
690 T---CAGAAGGAACCTCTGCAGGAAACCTGATGAGCTGCACCCCT---GGCGGAAGACAC 743
216 nProGluGlyThrSerAlaGlyThrProAspGluAlaAla-ProLeuAlaGluArgThrP 236
744 CTGACACGCTGAAAGCT---GGTGGAAAAACACCTGATGAGCTGCACCCCT---GGTGA 797
236 roAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAla-ProLeuValGlu 255
798 AAGACACCTGCACAGGCTGAAAGCT---GGTGGAAAAACACCTGATGAGCTGCATCT--- 851
256 ArgThrProAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAlaSerLeu 275
852 TGGTGGAGAACATCTGACAAATTCATGTTGGAGAAAGG---ACATCTGGAAGTTGGA 908
276 Val-GluGlyThrSerAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheGl 295
909 ACGTCAGC---AGAAAAACACCTAGGAAATACGATCC---TGCAAAAAACAATCTCAGA 962
295 uGln-SerAlaGluGluThrProArgGluIleThrSerProAlaLysGluThrSerGluL 315
963 AATTTAAG---GGCCAGCAAAAGAGACCTAGAGATCGCAGGGAGAA---AAGAAGAC 1016
315 ysPheThrTrpPro-AlaLysGlyArgProArgLysIleAlaTrpGluLysGluAsp 334
1017 ACACCTAGGGAATATGAGTCCCGAAAAAGAAACACTGAGAAA-----TTA 1061
335 ThrProArgGluIleMetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAla 354
1062 CGTGGCGCAAAAGGAAGACCTTAGGAAGACGATCGGGAAGAAAGAAA---CCCTGTAAA 1118
355 LysGlyArgProArgLysIle-----AlaTrpGluLysLysGluThrProValLys 371
1119 GATGATGCGT-GCAGAGTAACTACTAATAAATAAG---TTTGAAGAAAGAAATCT 1174
372 ThrGlyCysValAlaArgValThrSerAsnLysThrLysValLeuGluLysGlyArgSer 391
1175 AAGATGTT-GCATGTGCCAAAAG---AATCATCTCAAAAGCAAGTGCATGATCAGAGG 1230
392 LysMetIleAlaCysProThrLysGluSerThrLysAlaSerAlaAsnAspGlnArg 411
1231 ----TCCATCAGATCAAAACAGGGAAGATGAGATATTTCTTT-GATCTCGGCTCTCT 1286
412 PheProSerGluSerLysGlnGluAspGluGluTyrSerCysAspSerArg---Ser 430
1287 TTG-----AAGTTCTGCAAGATTCAAGTG---GPATACCTGAGTCTATATATAAAGT 1337
431 LeuPheGluSerSerAlaLysIleGlnValCysIleProGluSerIleTyrGlnLysVal 450
1338 AATGATATAATAAGAG---TAGAAAGCCTCCTAAAAGCCATCTGCTCTCAAGCGGCA 1394
451 MetGluIleAsnArgGluValGluGluProLysLysProSerAlaPheLysProAla 470
1395 TTGAAGCAAAA---CTCTTTCCAAATAAGCCTTTG-----AATGAAGAAATGAACA 1442
471 lIeGluMetGlnAsnSerValProAsnLysAlaPheGluLeuLysAsnGluGln----- 488
1443 AACATTGAGCAGATCCGTG-TTCCACCG---AATCCAAACAAAGACTATAGAAAT 1498
489 ThrLeuArgAlaAspProMetPheProGluSerLysGlnLysAspTyrGluGluAsn 508

```

| | | | |
|----|------|--|------|
| QY | 1499 | TC TTGGGATTC TAGAGTCTCTG---GAGACTGTTTACAGAAG---ATTGTGTTTACCAA | 1552 |
| Db | 509 | SerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLys | 528 |
| QY | 1553 | GGCTACACATCAAAAAGAAATAGATAAAAA---AATGGAAAA---TAGAAGAGTCCCTTAAT | 1606 |
| Db | 528 | SalAThrHisGlnLysGlu-IleAspLysIleAsnGlyLysLeuGluGluSerProAsnL | 548 |
| QY | 1607 | AAAGTGTCCTTCTGAAGCTACCTCGGAAT---GAAATTCATTATCCACTAAAGCCT--- | 1660 |
| Db | 548 | ysAspGlyLeuLeuLysAlaThrCys-GlyMetLysValSerIleProThrLysAlaLeu | 567 |
| QY | 1661 | AGAATTGAAGACATGCAAACTTTCAAAGCGAGCCTCC---GGGAAGCCATCTCCT--- | 1714 |
| Db | 568 | GluLeuLysAspMetGlnThrPheLysAlaGlu-ProProGlyLysProSerAlaPheG | 587 |
| QY | 1715 | GAGCCTCCACTGAAATCAAAAGTCTGTCCCAATTAAGCCTTGG-----AATGAAAA | 1765 |
| Db | 587 | uProAlaThrGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu-- | 606 |
| QY | 1766 | ATGAAAACATGGAGGCAGATGAAA---CTCCCATCAGATCCCAACAAGACTATAA | 1822 |
| Db | 607 | ---GlnThrTrpArgAlaAsp-GluLeuLeuProSerGluSerLysGlnLysAspTrpG | 625 |
| QY | 1823 | GA AAA---TTC TGGGATCTAGGACTCTCTG---TAGACTGTTCACAGAAGGATTG-TGT | 1875 |
| Db | 625 | LuGluAsnSerTrpAspThr-GluSerLeuCysGluThrValSerGlnLysAspValCys | 644 |
| QY | 1876 | TTACCAAG---CTGCCCTCAAAAGAAATAGATAAAATATGGAATTAAGAGGTT---1929 | 1929 |
| Db | 645 | LeuProLysAlaAlaHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGlySer | 664 |
| QY | 1930 | CCCTGTTTAAAGTGTCTTCTGAGCTTAAC---GGGAATCAAAAGTCTTATTCCAACTAAA | 1986 |
| Db | 665 | ProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLys | 684 |
| QY | 1987 | GCTTAGA---ATTGATGACATGCAAACTCAAAAG---CAGGCTCCCGAGAGCATCTGC | 2040 |
| Db | 685 | Ala-LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGlyLysProSerAl | 704 |
| QY | 2041 | CTTCGAGCCTCCATTGAAA---TCAAAAGTCTGCCAAATAAAC---CTTGGAAATTAAG | 2094 |
| Db | 704 | apHeGluProAlaIleGluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLysA | 724 |
| QY | 2095 | ATGAACAAAACATTGAGGCAGATGAGAA---CTCCCATCA---AATCCAAACAAGGACTA | 2148 |
| Db | 724 | snGluGlnThrLeuArgAlaAsp-GluIleLeuProSerGluSerLysGlnLysAspTrp | 743 |
| QY | 2149 | TGAAAAGTCTTGGGATTCGTGAGTCTCTG---TAGACTGTTTCCAGAAGGATGTG--- | 2203 |
| Db | 744 | GluGluSerSerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCy | 763 |
| QY | 2204 | -TTTACCAGGCTACACATCAAAAGAAATAGATAAAA---TAATGGAATAAGAGAGTC | 2259 |
| Db | 763 | sLeuProLysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSe | 783 |
| QY | 2260 | TCT-GATAATGAGGTTTTCTGAAGCTCCCT---GAGAATGAAAGTTCATTATCCACTAAA | 2315 |
| Db | 783 | rProAspAsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLy | 803 |
| QY | 2316 | GCCT---TGAATTGATGGCATGCAAACTTTCAAAGCAGGCCCTCCG---AGAGCCATCTG | 2369 |
| Db | 803 | SalALeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGlyLysProSerAl | 823 |
| QY | 2370 | CTTCGAGCCTGCATTGAAAT---GAAAAGTCTGTTCCAATAAAC---CTTGGAAATTAAG | 2423 |
| Db | 823 | apHeGluProAlaIle-GluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLys | 842 |
| QY | 2424 | AATGAACAACATTGAGACACATCAGAT-----TTCCTTCAGAAATCAAAACAAGAGG | 2477 |
| Db | 843 | AsnGluGlnThr---LeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLys | 861 |

| | | | |
|----|------|---|------|
| QY | 2478 | TTGAGAAAA---TTCTGGGATTCTGAAGTCT---CCGTGCACTGTGTTTCACAGAAAGGATGG | 2531 |
| Db | 862 | ValGluGlnAsnSerTrpAspSerGlu-SerLeuArgGluThrValSerGlnLysAsp-V | 881 |
| QY | 2532 | TGTTGTACC---CAGGCTACACACAAAAAGAAAGGATAAAAT---AGTGGAAAAATTAGAAG | 2585 |
| Db | 881 | alCysValProLysalaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluA | 901 |
| QY | 2586 | ATTCTAGTACGCTATCAAAAT---CTTGGTACAGTTTCAT---CTTGTGAAAGCGCAAGGAA | 2639 |
| Db | 901 | spSerThrSerLeuSerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluL | 921 |
| QY | 2640 | CTTCAAAAAGACACTGTGAACA-CGTACACGAAAAATGGAACAAT---GAAAAAGAGTTTT | 2695 |
| Db | 921 | euGlnLysAspHisCysGluGlnArgThrGlyLysMetGluGlnMetCysLysLysPheC | 941 |
| QY | 2696 | GTGTACTCAAAAAAGAAC---TGTCAGAACAAAAAGAAATAAATCACACTAGAGAA---CCA | 2749 |
| Db | 941 | ysValLeuLysLysLeuSerGluAlaLysGluIleLysSerGln-LeuGluAsnGln | 960 |
| QY | 2750 | AAAGCTTAATGGGAACAAGAG---TCTGCAGTGTAGATTCACTTAAACCAA---GAAAAG | 2803 |
| Db | 961 | LysValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGlu | 980 |
| QY | 2804 | AAAGACAGAGAAATGCCGATATATTAAAGAAAAA---TTAGCAAGAATTAG | 2851 |
| Db | 981 | LysArg-ArgAsnAlaAspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIle-G | 1000 |
| QY | 2852 | AAGAAATCGAGCAGCAGTAGAAGAGTTAGAGTGAACAACAACACTTGAAAG---GCTCTC | 2908 |
| Db | 1000 | luGluGlnHisArg-----LysgluLeuGluValLysGlnGlnLeu-GluGlnAlaLeu | 1017 |
| QY | 2909 | AGATACAAG---ATAAGAAATTGAAGGTGTAGAAGTAATTTGAATAGTTTCTCA---AC | 2962 |
| Db | 1018 | ArgIleGlnAspIleGluLeuLysSerValGluSerAsnLeuAsnGln-ValSerHisThr | 1037 |
| QY | 2963 | TCATGAAATGAAA---ATTATTCTTACATGAAATTCGATCTGAAAAAGAAATGCCA-- | 3017 |
| Db | 1037 | rHisGluAsnGluAsnTrpLeuLeuHisGluAsnCysMetLeuLysLysGluIleAlaMe | 1057 |
| QY | 3018 | -TGCAAAACTGGAATGCCACACTA---ACACCATACTCAGGAAAAAGAAAAATAATC-T | 3072 |
| Db | 1057 | tLeuLysLeuGluIlealaThrLeuLysHis-GlnTrpGlnGluLysGluAsnLysTrp | 1077 |
| QY | 3073 | TTGAGGACTTAAGATTTTAAAGA---AAAGATCTGAACCTTAG-ATCACCTAAAACTGA | 3128 |
| Db | 1077 | heGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuGlnMetThrLeuLysLeuL | 1097 |
| QY | 3129 | AAAGGAATCATTTACTAAAAG---GGATCTCAATATGTGGCACCT---AAAGTTCTGATA | 3182 |
| Db | 1097 | ysGluGluSerLeuThrLysArgAlaSerGlnTrpSerGly-GlnLeuLysValLeuIle | 1116 |
| QY | 3183 | GCTGAGAAAACAATGCTCATTTCTAAATTGAGAAAAACACAAAGAAAA---CTAGA | 3236 |
| Db | 1117 | AlaGlu-AsnThrMetLeuThrSerLysLysLeuLysGlnAspLysGluIleLeuGlu | 1136 |
| QY | 3237 | GGCAGAAATTGATA-CACCATCCTGACTGGCTTCGCTGTACAAGC-----CA | 3283 |
| Db | 1136 | uAlaGluIleGluSerHisHisPro-----ArgLeuAlaSerAlaValGlnAspHi | 1153 |
| QY | 3284 | TGATCAATTGTGACATCAACAGAAAAAGTAAGA---ACCTGCTCCACATTCAG---CAGAT | 3337 |
| Db | 1153 | sAspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl | 1173 |
| QY | 3338 | GCTGTTTGCAGAGAAATGAAATCTTCATGTG---AGAGTACGATATTAAACAATGAGTCT | 3394 |
| Db | 1173 | aCysLeuGlnArgLysMetAsnValAspValSerSerThrIleTrp-AsnAsnGluValL | 1193 |
| QY | 3395 | CCATCACCACCTTC---TGAGCTCAAGGAATCCAAAACCTTAAATATCT---CAAT | 3448 |
| Db | 1193 | euHisGlnProLeuSerGluAlaGlnArgLysSerLysSerLeuLysIleAsnLeuAsnTr | 1213 |
| QY | 3449 | TAGCAGAGAGATGCTCAAGAGA---AATACATTTGGTTTCAGAACATCAACAAGAGA---CA | 3502 |

```

Db 1213 yrAlaGlyAsp-AlaLeuA-GluAsnThrLeuValSerGluHis-AlaGlnArgAspGI 1232
QY 3503 ACGTGAACACAGTGTCAAT---GAAGGAAGCGAACACATGTATCAAAAC-----GACAA 3553
Db 1332 nArgGluThrGlnCysGlnMetLysGluAlaGluHisMetTyrGlnAsnGluGlnAspAs 1252
QY 3554 GATAATGGAACAAACACTGAACAGCAGAGTCTTCTAGTCAGAAATTTATTTCACTACAG 3613
Db 1252 nValAsnLysHisThrGluGlnGlnGluSerLeu-AspGlnLysLeuPheGlnLeuGlns 1272
QY 3614 CAAAAATGTGCTTCACAGCAAT---AGTCATGCACATAGAAAGCTCACAACAAAAG 3669
Db 1272 erLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaHisLys-----LysA 1288
QY 3670 CAGATAACAAT-----GATATTCATTTCTTG---AGAGGAATATGCAACAC 3711
Db 1288 laAspAsnLysSerLysileThrileAspilleHisPheLeuGluArgLysMetGlnHisH 1308
QY 3712 ATCTCCTAAAGAGAAAAGAGGAGATATT---AATTACATACCATTTAAAGACCGTAT 3768
Db 1308 isLeuLeuLysGluLysAsnGlu-GluilePheAsnTyrAsnAsnHisLeuLysAsnArg 1327
QY 3769 ATT---CAATATGAAAAAGAAAA 3789
Db 1328 ilLeyrGlnTyrGluLysGluLys 1335

RESULT 3
US-10-076-622-565
; Sequence 565, Application US/10076622
; Publication No. US20030023036A1
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C11
; CURRENT APPLICATION NUMBER: US/10/076,622
; CURRENT FILING DATE: 2002-02-13
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 565
; LENGTH: 1341
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-076-622-565

Alignment Scores:
Pred. No.: 8,89e-110 Length: 1341
Score: 1234.00 Matches: 679
Percent Similarity: 59.18% Conservative: 146
Best Local Similarity: 48.71% Mismatches: 376
Query Match: 17.41% Indels: 207
Db: 14 Gaps: 115

US-09-602-362E-22 (1-4115) x US-10-076-622-565 (1-1341)
QY 96 CAAAGGGAAGAGACATCAACTTAATATACAAGAG---CCGAGAAGAGACTGCTCTAAC 152
Db 3 LysArgLysLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu-- 21
QY 153 TGGGCTGTGCTCAATGGCTC-----GAGGAAGTAGTAACATTTC---GGTA 194
Db 22 -----HisTrpAlaCysValAsnGlyHisGlnGluValValThr-PheLeuVala 38
QY 195 GACAGAAGTGCAGTGACGCTCTTG---TGCGCAACACGACACCTCTGATGAAGG-T 250
Db 38 spArgLysCysGlnLeuAspValLeuAspGlyGluHisArgThrProLeuMetLysAlaL 58
QY 251 TACAATGGCACAGGAGGCTTTGTC---AAATATTGATAGATTC---GGTCCGATATAATC 306
Db 58 euGlnCysHisGlnGlnAlaCysAlaAsnIleLeuileAspSerGlyAlaAspIleAsnL 78
```

```

QY 307 TCGTG---ATGTGTATGCAACATCGCTTCCATTATGCGTTTATAGTGAGATTTTGTCACT- 362
Db 78 euValAspValTyrGlyAsnMetAlaLeuHisTyrAlaValTyrSerGluileLeuSerV 98
QY 363 --GGTGGCAAACTGTGTCT-CATGCTGCAGTATCGAAGTGC---CAACAAGCTGCCCTCA 416
Db 98 alValAlaLysLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeuT 118
QY 417 CACCACTTTTACTACCAT---AACGAAGAAGTGCAGCAATTTGTGAATTTT---GCTCAT 470
Db 118 hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuLeuIle 137
QY 471 AAAAATGCAATCGAATGCGTTA---ATAAGTTAAATGCACACCCTCATGCTGCTGTAT 527
Db 138 LysAsnAlaAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAlaVa 157
QY 528 G---TCTGATCATCAGAGATAGTTGCAATGCTTCTTACCAAAA---TGTGACGCTCTTT-- 579
Db 157 lCysHisGlySerSerGluIleValGlyMetLeuLeuGlnGlnAsnValAspValPheAl 177
QY 580 -CTGCAGATATAGTGAAGTAACTGCAGAACAT---ATGCTGTACTGTGGATTCTTCTACA 635
Db 177 aAlaAspIle-CysGlyValThrAlaGluHisTyrAlaValThrCysGlyPheHisIle 197
QY 636 TTCATAACAAA---TTATGAATATATACGAAAATTTATCAAAATCA---TCAATATACCAA 689
Db 197 leHisGluGlnIleMet-GluTyrIleArgLysLeuSer-LysAsnHisGlnAsnThrAs 216
QY 690 T---CAGAAGAACTCTGCAGAAACCTGTAGAGCTGCACCT---GGCGGAAGACAC 743
Db 216 nProGluGlyThrSerAlaGlyThrProAspGluAlaAla-ProLeuAlaGluArgThrP 236
QY 744 CTGACACGCTGAAAGCT---GGTGGAAAAACACCTGTAGAGCTGCACCT---GGTGGGA 797
Db 236 roAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAla-ProLeuValGlu 255
QY 798 AAGACACCTGACAGCTGAAAGCT---GGTGGAAAAACACCTGTAGAGCTGCACCT--- 851
Db 256 ArgThrProAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAlaSerLeu 275
QY 852 TGTGTGAGGAACATCTGCACAAATTTCAATGTTGGAGAAAG---ACATCTGGAAGTTTCCA 908
Db 276 Val-GluGlyThrSerAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheGl 295
QY 909 ACCTCAGC---AGAAAAACACCTAGAAATATACAGTCC---TGCAAAAAACATCTGAGA 962
Db 295 uGln-SerAlaGluThrProArgGluIleThrSerProAlaLysGluThrSerGluL 315
QY 963 AATTTACG---GGCCAGCAAGGAAGACCTAGAAGATCGAGGAGAA---AAAGAAGAC 1016
Db 315 yspPheThrTrpPro-AlaLysGlyArgProArgLysIleAlaTrpGluLysLysGluAsp 334
QY 1017 ACACCTTAGGGAATATGAGTCCGAAAAAGAAACATCTGAGAAA-----TTA 1061
Db 335 ThrProArgGluIleMetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAla 354
QY 1062 CGTGGCAGAAAAGGAGACCTAGGAAGCGCATGGAGAAAAAGAAA---CCCTGTAAA 1118
Db 355 LysGlyArgProArgLysIle-----AlaTrpGluLysLysGluThrProValLys 371
QY 1119 GATGATGCGCT-GCAAGAGTAACTCTATAAACTAAAG---TTTGTAAAAAGGAATCT 1174
Db 372 ThrGlyCysValAlaArgValThrSerAsnLysThrLysValLeuGluLysGlyArgSer 391
QY 1175 AAGATGTT-GCATGTCCACAAAG---AATCATCTACAAGCAAGTGCATCATCAGAGG 1230
Db 392 LysMetIleAlaCysProThrLysGluSerThrLysAlaSerAlaAsnAspGlnArg 411
QY 1231 ---TCCCATCAGATCCAAAGGGAAGATGAAGATATTTCTTT-GATTCGGGTCTCT 1286
Db 412 PheProSerGluSerLysGlnGluAspGluGluTyrSerCysAspSerArg---Ser 430
```


QY 1287 TTG-----AAAGTTCTGCAAGATTCAAGTG---GTATACCTGAGTCTATATATAAAAGT 1337
Db |||||
431 LeuPheGluSerSerAlaLysIleGlnValCysIleProGluSerIleTyrGlnLysVal 450
QY 1338 AATGAGATAAATAAGAG---TAGAAGCCTCTTAAAGCCATCTGCTTCAAGCCGCCA 1394
Db |||||
451 MetGluIleAsnArgGluValGluGluProProLysLysProSerAlaPheLysProAla 470
QY 1395 TTGAAGCAAAA---CTCTTTCCAAATAAGCCTTTG-----AATGAAGAAATGAACA 1442
Db |||||
471 IleGluMetGlnAsnSerValProAsnLysAlaPheGluLeuLysAsnGluGln----- 498
QY 1443 AACATTAGACAGATCCGTG-TTCCACCG---AATCCAACAAGAGACTATAGAANAAT 1498
Db |||||
489 ThrLeuArgAlaAspProMetPheProGluSerLysGlnLysAspTyrGluGluAsn 508
QY 1499 TCTTGGGATTTAGAGTCTCTG---GAGACTGTTTACAGAGG---ATTGTGTTTACCA 1552
Db |||||
509 SerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLy 528
QY 1553 GGCTACACATCAAAAAGANTAGATAAAA---AATGAAAA---TAGAAGAGTCCCTPAAT 1606
Db |||||
528 sAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsnL 548
QY 1607 AAAGTGTCTTCTGAAGCTACCTCGGAAT---GAAATTTCTATTCACCTAAAGCCT--- 1660
Db |||||
548 ysAspGlyLeuLeuLysAlaThrCys-GlyMetLysValSerIleProThrLysAlaLeu 567
QY 1661 AGAATTGAAGACATGCAAACTTTCAAGCGAGCCTCC---GGGAAGCCTCTGCT---TC 1714
Db |||||
568 GluLeuLysAspMetGlnThrPheLysAlaGlu-ProProGlyLysProSerAlaPheG 587
QY 1715 GAGCCTCCACTGAATCAAAAGTCTGCCAAATAAGCCTTGG-----AATGAAAA 1765
Db |||||
587 uProAlaThrGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu-- 606
QY 1766 ATGAAAAACATGGAGGAGATGAGAA---CTCCCATCAGAAATCAAAAGAGACTATAA 1822
Db |||||
607 ---GlnThrTrpArgAlaAsp-GluIleLeuProSerGluSerLysGlnLysAspTyrG 625
QY 1823 GAAAA---TTCTGGGATCTGGAGTCTCTG---TAGACTGTTTCAAGAGGATG-TGT 1875
Db |||||
625 luGluAsnSerTrpAspThr-GluSerLeuCysGluThrValSerGlnLysAspValCys 644
QY 1876 TTACCAAGG---CTGCGCTCAAAAAGAAATAGATAAAATATGAAATATAGAGGCT--- 1929
Db |||||
645 LeuProLysAlaAlaHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGlySer 664
QY 1930 CCTGTGTTAAAGTGTCTTCTGAGGCTAACT---GGGAATGAAAAGTCTCTATTTCCAATAA 1986
Db |||||
665 ProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLys 684
QY 1987 GCCTAGA---ATTGATGACATGCMAACTTCAAAG---CAGGCTCCGAGAGCCTCTGC 2040
Db |||||
685 Ala-LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl 704
QY 2041 CTTGAGCCTCCATTAAGAA---TCAAAAGTCTGTCCAAATAAAAC---CTTGGAAATTAAGA 2094
Db |||||
704 aPheGluProAlaIleGluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLysA 724
QY 2095 ATGACAAAAATTGAGCAGATAGAA---CTCCCATCA---AATCCAACAAGAGACTA 2148
Db |||||
724 snGluGlnThrLeuArgAlaAsp-GluIleLeuProSerGluSerLysGlnLysAspTyr 743
QY 2149 TGAGAAAGTCTTGGGATTTGAGTCTCTG---TAGACTGTTTCCAGAGGATGCTG--- 2203
Db |||||
744 GluGluSerSerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCy 763
QY 2204 -TTTACCCAGGCTACATCAAAAAGAAATAGATAAAA---TAATGAAAATAGAGAGTCT 2259
Db |||||
763 sLeuProLysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSe 783
QY 2260 TCT-GATAATGAGGTTTCTGAAGGCTCCCT---GAGATGAAGATTCATTCCACTAA 2315

Db |||||
783 rProAspAsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLy 803
QY 2316 GCCT---TGAATTGATGGCATGCAAACTTTCAAGAGCAGCCTCCCG---AGAGCCATCTG 2369
Db |||||
803 sAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAl 823
QY 2370 CTTGAGCCTGCAATTAATAAT---GAAAAGTCTGTTCCTCAATAAAT---CTTGGAAATTAAG 2423
Db |||||
823 aPheGluProAlaIle-GluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLys 842
QY 2424 AATGAACAACATTGACAGACATCAGAT-----TTCCCTTCAGAAATCAAAAACAAAGAG 2477
Db |||||
843 AsnGluGlnThr---LeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLys 861
QY 2478 TTGAGAAAA---TTCTGGGATTTCTGAAGTCT---CCGTGAGCTGTTTTCACAGAGATG 2531
Db |||||
862 ValGluGluAsnSerTrpAspSerGlu-SerLeuArgGluThrValSerGlnLysAsp-V 881
QY 2532 TGTGTACC---CAGGCTACACAAAAAGAAAGATATAAAT---AGTGGAAATTTAGAG 2585
Db |||||
881 alCysValProLysAlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluA 901
QY 2586 ATTCTACTACCTTCAAAAT---CTTGTGACTAGTTTCAT---CTTGTGAAGCAAGGNA 2639
Db |||||
901 sPserThrSerLeuSerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluL 921
QY 2640 CTTCAAAAAGACACTGTGAAACA-CGTACAGGAAAAATGGAACAAT---GAAAAGAGTTT 2695
Db |||||
921 euGlnLysAspHisCysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheC 941
QY 2696 GTGTACTGAAAAGAAC---TGTCAACAAGAAATAAATACACAGTAGAGAA---CCA 2749
Db |||||
941 ysValLeuLysLysLysLeuSerGluAlaLysGluIleLysSerGln-LeuGluAsnGln 960
QY 2750 AAAAGTTAAATGGAAACAAGAG---TCTGAGTGTAGTACTTAAACCAA---GAAAAG 2803
Db |||||
961 LysValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGlu 980
QY 2804 AGAAGAGAAATGCCGATATATTAAAGAAAAA-----TTAGGAAGAAATTAG 2851
Db |||||
981 LysA-g-ArgAsnAlaAspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIle-G 1000
QY 2852 AAGAATCGAGACACCATAGAAAGATTAGAGTGAACACAACTTGAAG---CCTCTC 2908
Db |||||
1000 luGluGlnHisArg-----LysGluLeuGluValLysGlnGlnLeu-GluGlnAlaLeu 1017
QY 2909 AGATACAAG---ATAAGAAATTGAAGTGTAGAAAGTAATTTGAATAGGTTTCTCA---AC 2962
Db |||||
1018 ArgIleGlnAspIleGluLeuLysSerValGluSerAsnLeuAsnGln-ValSerHisTh 1037
QY 2963 TCATGAATGAAA---ATTATCTTACATGAATTCATGTTGATAAAGAAAGTAATGCCA--- 3017
Db |||||
1037 hisGluAsnGluAsnTyrLeuLeuHisGluAsnCysMetLeuLysLysGluIleAlaMe 1057
QY 3018 -TGCAAAATCGAAATAGCCACACTA---AACACCAATCCAGAAAGGAAATAAATC-T 3072
Db |||||
1057 tleuLysLeuGluIleAlaThrLeuLysHis-GlnTyrGlnGluLysGluAsnLysTyrP 1077
QY 3073 TTGAGAGCTTAAGATTTTAAAGA---AAAGATGCTGAACCTTAG-ATGACCTTAAACTGA 3128
Db |||||
1077 heGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuGlnMetThrLeuLysLeuL 1097
QY 3129 AAAGGAATCATTTACTTAAAG---GGATCTCAATATGTGGCAGCT---AAAGTTCTGATA 3182
Db |||||
1097 ysGluGluSerLeuThrLysArgAlaSerGlnTyrSerGly-GlnLeuLysValleuIle 1116
QY 3183 GCTGAGAAACAATGCTCTATTCTAAATTGAGAAAAACAGACAAAGAAAA-----CTAGA 3236
Db |||||
1117 AlaGlu-AsnThrMetLeuThrSerLysLeuLysGluLysGlnAspLysGluIleLeuG 1136
QY 3237 GCGAAGATTCGATA-CACCATCTGACTGCTGCTCGCTGTACACAGC-----CA 3283

1136 ualaGluileGluSerHisHisPro-----ArgLeuAlaSerAlaValGlnAspHi 1153
3284 TGATCAATTGACATCAAGAAAAAGTAAGA--ACCTGCTTCCACATTGAG---GAGAT 3337
1153 sAspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl 1173
3338 GCTGTTGCAAGAAATGAATGTTGATG---AGATGACGATTAATCAACATGAGTCT 3394
1173 aCysLeuGlnArgLysMetAsnValAspValSerThrIleTyr-AsnAsnGluValL 1193
3395 CCATCACCACCTTC---TGACCTCAAGGAATCAAAACCTAAATAATATCT---CAAT 3448
1193 euHisGlnProLeuSerGluAlaGlnArgLysSerLysSerLeuLysIleAsnLeuAsnT 1213
3449 TAGCAGGAGATGTTCAAGAGA---AATACATGTTTTCAGAACATCAACAAGAGA---CA 3502
1213 yAlaGlyAsp-AlaLeuArgGluAsnThrLeuValSerGluHis-AlaGlnArgAspGl 1232
3503 ACCTGAAACACAGTGTCAAT---GAAGGAAGGAAACACATGTATCAAAAC-----GACAA 3553
1232 nArgGluThrGlnCysGlnMetLysGluAlaGluHisMetTyrGlnAsnGluGlnAspAs 1252
3554 GATATGGAACAAACACCTCAACAGCAGAGTCTTAGTCAGAAATATTCAACTACAAG 3613
1252 nValAsnLysHisThrGluGlnGlnGluSerLeu-AspGlnLysLeuPheGlnLeuGlnS 1272
3614 CAAAATTTGCTTCACAGCAATT---AGTCATGCACATAGAAAGCTGACACAAAG 3669
1272 erLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaHisLys-----LysA 1288
3670 CAGATAACAAT-----GATATTCTATTTCTTG---AGAGGAATGCAACAC 3711
1288 laAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLysMetGlnHisH 1308
3712 ATCTCCTAAAAGAGAAAGAGGATATT---AATTACATACATTTAAACCCGTAT 3768
1308 isLeuLeuLysGluLysAsnGlu-GluIlePheAsnTyrAsnAsnHisLeuLysAsnArg 1327
3769 ATT---CAATATGAAAAAGAGAAA 3789
1328 lietyrGlnTyrGluLysGluLys 1335
RESULT 4
US-10-177-293-334
; Sequence 334, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavaru, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; CURRENT APPLICATION NUMBER: US/10/177,293
; FILE REFERENCE: MRI-038
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572

PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 334
LENGTH: 1341
TYPE: PRT
ORGANISM: Homo sapiens
US-10-177-293-334
Alignment Scores:
Pred. No.: 8,89e-110 Length: 1341
Score: 1234.00 Matches: 679
Percent Similarity: 59.18% Conservative: 146
Best Local Similarity: 48.71% Mismatches: 376
Query Match: 17.41% Indels: 207
DB: 14 Gaps: 115
US-09-602-362E-22 (1-4115) x US-10-177-293-334 (1-1341)
QY 96 CAAAGGGAAGACATCAACCTTAATATACAAGAG---CCGAGAAGAGACTGCTCTAAC 152
DB 3 LysArgLysLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu-- 21
QY 153 TGGGCTGTCTCAATGCGCT-----GAGGAAGTAGTAACATTC---GGTA 194
DB 22 -----HisTrpAlaCysValAsnGlnHisGluValValThr-PheLeuVala 38
QY 195 GACAGAGTGCAGCTGACGCTCTTG---TGGCGAACAACGACACCTCTGTATGAGGC-T 250
DB 38 spArgLysCysGlnLeuAspValLeuAspGlyGluHisArgThrProLeuMetLysAlaL 58
QY 251 TACAATGCCACAGAGGCTTTCG---AAATATTGATGATTC-GTGCCGATATAATC 306
DB 58 euGlnCysHisGlnGluAlaCysAlaAsnIleLeuIleAspSerGlyAlaAspIleAsnL 78
QY 307 TCGTG---ATGTGTATGCAACATGCTTCCATTATGCGTTTATAGTGGATTTTTCAGT- 362
DB 78 euValAspValTyrGlyAsnMetAlaLeuHisTyrAlaValTyrSerGluIleLeuSerV 98
QY 363 --GGTGGCAAACTGTGTC-CATGCTGCAGATGCAAGTGC---CAACAAGCTCCCTCA 416
DB 98 alValAlaLysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeuT 118
QY 417 CACCCTTTTACTACCAT---AACGAAAGAAGTGCATTCGGAATTTT---GCTGAT 470
DB 118 hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuLeu 137
QY 471 AAAAATGCAAAATCGCAATCGTTA---ATAAGTTAAATGCACACCTCATGCTGTGAT 527
DB 138 LysAsnAlaAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAlaVa 157
QY 528 G---TCTGGATCATCAGATAGTTGCGATCTCTTAGCAAAA---TGTGACGCTCTT-- 579
DB 157 iCysHisGlySerSerGluIleValGlyMetLeuLeuGlnGlnAsnValAspValPheAl 177
QY 580 -GTGCAGATATAGTGAGTAACCTGCAAGAACAT---ATGCTGTACTGTGGATTTCTCACA 635
DB 177 aaAlaAspIle-CysGlyValThrAlaGluHisTyrAlaValThrCysGlyPheHisIle 197
QY 636 TTCATAACAAA---TTATGAATATATACGAAATATCAAAATCA---TCAAAATACCAA 689
DB 197 leHisGluGlnIleMet-GlutryIleArgLysLeuSer-LysAsnHisGlnAsnThrAs 216
QY 690 T---CAGAAGGAACCTCTGCAGAAACCTGATGAGGTGACACCT---GGCGGAAAGACAC 743

Dbb 216 nProGluGlyThrSerAlaGlyThrProAspGluAlaAla-ProLeuAlaGluArgThrP 236
Qy 744 CTGACAGCTGAAAGCT---GGTGGAAAAACACCTGATGAGGTGCACCCCT---GGTGGGA 797
Dbb 236 roAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAla-ProLeuValGlu 255
Qy 798 AAGACACTGACAGCTGAAAGCT---GGTGGAAAAACACCTGATGAGGTGCATCT--- 851
Dbb 256 ArgThrProAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAlaSerLeu 275
Qy 852 TGGTGGAGGACACTCTGACAAATCAATGTTGGAGAAAG---ACATCTGGAAGTTCGA 908
Dbb 276 Val-GluGlyThrSerAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheG 295
Qy 909 ACCTCAGC---AGAAAAACACTAGGAAATPACGATCC---TGCAAAAAACACTCTGAGA 962
Dbb 295 uGln-SerAlaGluGluThrProArgGluIleThrSerProAlaLysGluThrSerGlu 315
Qy 963 AATTTCAG---GGCCACCAAGGAGACCTAGAGATCGCAGGAGAA---AAGAGAGAC 1016
Dbb 315 ysPheThrTrpPro-AlaLysGlyArgProArgLysIleAlaTrpGluLysLysGluAsp 334
Qy 1017 ACACCTAGGGAATPATGAGTCCCGAAAAAGAAACACTGAGAAA-----TTA 1061
Dbb 335 ThrProArgGluIleMetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAla 354
Qy 1062 CGTGGGAGAAAAGGAGACCTAGGAGAGCGATGGGAGAAAAGAAA---CCCTGTAAA 1118
Dbb 355 LysGlyArgProArgLysIle-----AlaTrpGluLysLysGluThrProValLys 371
Qy 1119 GATGGATCGCT-GCAAGAGTAACATCTAATAAACTAAAG---TTTTCAAAAAGGAATCT 1174
Dbb 372 ThrGlyCysValAlaAlaArgValThrSerAsnLysThrLysValLeuGluLysGlyArgSer 391
Qy 1175 AAGATGTT-GCATCTCCAAAAAG---AATCATCTCAAAAGCAAGTGCATGATCAGAG 1230
Dbb 392 LysMetIleAlaCysProThrLysGluSerSerThrLysAlaSerAlaAsnAspGlnArg 411
Qy 1231 ---TCCCATCAGATCCAAAACAAGGAAGAGTGAAGATATTCTTT-GATTCTCGGGTCTCT 1286
Dbb 412 PheProSerGluSerLysGlnGluGluAspGluGluTrpSerCysAspSerArg---Ser 430
Qy 1287 TTG-----AAGTCTCCAGATTCAAGTG---GTATACCTGAGTCTATATATAAAAGT 1337
Dbb 431 LeuPheGluSerSerAlaLysIleGlnValCysIleProGluSerIleTyrGlnLysVal 450
Qy 1338 AATGAGTAATAAGAG---TAGAAGCCTCTTAAGACCATCTGCCTCAAGCCGCCA 1394
Dbb 451 MetGluIleAsnArgGluValGluGluProProLysLysProSerAlaPheLysProAla 470
Qy 1395 TTGAAAGCAAAA---CTCTTTCCAAATAAGCCTTTG-----AATGAAGATGAACA 1442
Dbb 471 IleGluMetGlnAsnSerValProAsnLysAlaPheGluLeuLysAsnGluGln----- 488
Qy 1443 AACATTGAGCAGATCCGTG-TTCCCAACG---AATCCAACAAAGGACTATAGAATAAT 1498
Dbb 489 ThrLeuArgAlaAspProMetPheProGluSerLysGlnLysAspTyrGluGluAsn 508
Qy 1499 TCTGGGATTCTAGACTCTCTG---GAGACTGTTTACAGAAAG---ATTGTGTTACCAA 1552
Dbb 509 SerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProly 528
Qy 1553 GGCTACACATCAAAAAGATAGATAAAA---AATGGAAAA---TAGAAGAGTCCCTAAT 1606
Dbb 528 salAThrHisGlnLysGlu-IleAspLysIleAsnGlyLysLeuGluGluGluSerProAsnL 548
Qy 1607 AAGTGTGCTCTCAAGGCTACCTCGGAAT---GAAATTCTATTCCACTAAGGCCT--- 1660
Dbb 548 ysAspGlyLeuLeuLysAlaThrCys-GlyMetLysValSerIleProThrLysAlaLeu 567
Qy 1661 AGAATTGAAGACATGCAAACTTTCAAAAGCGAGCGCTCC---GGGAAGCCATCTGCT---TC 1714
Dbb 568 GluLeuLysAspMetGlnThrPheLysAlaGlu-ProProGlyLysProSerAlaPheG 597

Qy 1715 GAGCCTCCACTGAAATCAAAAGTCTGTCCCAATTAAGCCTTGG-----AATGAAAA 1765
Dbb 587 uProAlaThrGluMetGlnLysSerValProAsnLysAlaLeuGluLysAsnGlu--- 606
Qy 1766 ATGAAAAACATCGAGCGCAGATGAGAA---CTCCCATCAGATCCAAAGCAAGGACTATAA 1822
Dbb 607 ---GlnThrTrpArgAlaAsp-GluIleLeuProSerGluSerLysGlnLysAspTyrG 625
Qy 1823 GAAAA---TTCTGGGACTACTGAGTCTCTG---TAGACTGTTTTCACAGAGGATTG-TGT 1875
Dbb 625 luGluAsnSerTrpAspThr-GluSerLeuCysGluThrValSerGlnLysAspValCys 644
Qy 1876 TTACCAAGG---CTCGGCTCAAAAAGATAGATAAAATAATGGAAAAATTAGAGGGT--- 1929
Dbb 645 LeuProLysAlaAlaHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGlySer 664
Qy 1930 CCTGTAAAGTGGTCTTCTGAGGCTAACT---GGGAATGAAGTCTTATTCCTCAACTAAA 1986
Dbb 665 ProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLys 684
Qy 1987 GCTAGA---ATTGATGACATGCAAACTTCAAG---CAGGCTCCCGAGAGCCATCTGCG 2040
Dbb 685 Ala-LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl 704
Qy 2041 CTTTCGAGCTCCATTGAAA---TCAAAAGTCTGTCCCAATAAAC---CTTGAATTAAGA 2094
Dbb 704 aPheGluProAlaIleGluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLysA 724
Qy 2095 ATGAACAAACATTGAGCAGATGAGAA---CTCCCATCA---AATCCAAACAAAGAGACTA 2148
Dbb 724 snGluGlnThrLeuArgAlaAsp-GluIleLeuProSerGluSerLysGlnLysAspTyr 743
Qy 2149 TGAGAAAGTCTTGGGATCTGGAGTCTCTG---TAGACTGTTTTCACAGAGGATGCTG--- 2203
Dbb 744 GluGluSerSerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCy 763
Qy 2204 -TTTACCAGGCTCACATCAATCAAAAGATAGATAAAA---TAATGGAATAAGAGAGTC 2259
Dbb 763 slenProLysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSe 783
Qy 2260 TCT-GATTAATGAGGTTTCTGAGAGCTCCCT---GAGAAATGAAGTCTTATTCCTCAATAA 2315
Dbb 783 rProAspAsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrly 803
Qy 2316 GCCT---TGAATTGATGCGATGCAAACTTCAAGCAGGCTCCCG---AGAGCCATCTG 2369
Dbb 803 salalLeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl 823
Qy 2370 CTTGAGGCTCGATTGAAAT---GAAAAGTCTGTTCCTCAATAAAC---CTTGAATTAAG 2423
Dbb 823 aPheGluProAlaIle-GluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLys 842
Qy 2424 AATGAACAACTGAGCAGATCAGAT-----TTCCCTTCAGATCAAAACAAAGAGG 2477
Dbb 843 AsnGluGlnThr---LeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLys 861
Qy 2478 TTGAGAAAA---TTCTGGGATCTCAAGTCT---CCGTGGACTGTTTTCACAGAGGATCG 2531
Dbb 862 ValGluGluAsnSerTrpAspSerGlu-SerLeuArgGluThrValSerGlnLysAsp-V 881
Qy 2532 TGTGTACC---CAGCTTCACAAAAAAGAGGATAAAAT---AGTGGAAAAATTAGAG 2585
Dbb 881 aiCysValProLysAlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluA 901
Qy 2586 ATTCTACTAGCTTATCAAAAT---CTTGTACAGTTTCAT---CTTGTGAAGCAAGGAGAA 2639
Dbb 901 spSerThrSerLeuSerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluL 921
Qy 2640 CTTCAAAAAGACACTGTGAACA-CGTACAGGAAAATGGAACAAT---GAAAAGAGTTT 2695
Dbb 921 euGlnLysAspHisCysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheC 941

```
QY 2696 GTGTACTGAAAGAAC---TGTGACACAAAGAAATAATCACAGTAGAGAA---CCA 2749
Db 941 ysvallLeuLysLysLeuSerGluAlaLysGluLysSerGln-LeuGluAsnGln 960
QY 2750 AAAGTTAAATGGACACAGAG---TCTGCAGTGTAGATTGACCTTAAACCAA---GAAAG 2803
Db 961 LysValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGlu 980
QY 2804 ACAAGAGAAATGCCGATATATTAAAGAAAAA---TTAGGAGAATTAG 2851
Db 981 LysArg-ArgAsnAlaAspIleLeuAsnGluLysIleArgGluLeuGlyArgIle-G 1000
QY 2852 AAGATCGAGAGCAGCAGTAGAAGAGTTAGAGTGAACACAACTTGAAG---GCTCTC 2908
Db 1000 luGluGlnHisArg-----LysGluLeuGluValLysGlnGlnLeu-GluGlnAlaLeu 1017
QY 2909 AGATACAAAG---ATAAGAAATTGAAGGTGTAGAAAGTAAATGAGTTCTCA---AC 2962
Db 1018 ArgIleGlnAspIleGluLeuLysSerValGluSerAsnLeuAsnGln-ValSerHisTh 1037
QY 2963 TCATGAAATGAAA---ATTATCTTACATGAAATGATGTGTAAGAAAGAAATGCCA--- 3017
Db 1037 rHisGluAsnGluAsnTyrLeuLeuHisGluAsnCysMetLeuLysLysGluIleAlaMe 1057
QY 3018 -TGCAAACTGGAATAGCAGCAGCTA---AACACCAATCCAGAAAGAAATAAATC-T 3072
Db 1057 tLeuLysLeuGluIleAlaThrLeuLysHis-GlnTyrGlnGluLysGluAsnLysTyrP 1077
QY 3073 TTGAGGACTTAAGATTTTAAAGA---AAAGATGCTGAACCTTAG-ATGACCTTAAACTGA 3128
Db 1077 heGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuGlnMetThrLeuLysLeuL 1097
QY 3129 AAAGAAATCATCTAAAG---GGATCTCAATATGTCGAGCT---AAAGTCTGATA 3182
Db 1097 ysGluGluSerLeuThrLysArgAlaSerGlnTyrSerGly-GlnLeuLysValLeuIle 1116
QY 3183 GGTGAGAAACATGCTCATCTTAATTCAGGAAAGAAACAGACAAAGAAAA---CTAGA 3236
Db 1117 AlaGlu-AsnThrMetLeuThrSerLysLeuLysGluLysGlnAspLysGluIleLeuGl 1136
QY 3237 GCGAAGATTTGAATA-CACCATCTGACTGCTTCGCTGTCACAGC-----CA 3283
Db 1136 uAlaGluIleGluSerHisPro-----ArgLeuAlaSerAlaValGlnAspHi 1153
QY 3284 TGATCAATTTGACATCAAGAAAAAGTAAGA---ACCTGCTTCACATTCAG---GAGAT 3337
Db 1153 sAspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl 1173
QY 3338 GCTGTTTCCAAAGAAATGAATGTTGATGTG---AGAGTAGATATTAACAATGAGTGTCT 3394
Db 1173 aCysLeuGlnArgLysMetAsnValAspValSerSerThrIleTyr-AsnAsnGluValL 1193
QY 3395 CCATCACCACITTC---TGACTCAAGGAAATCCAAACCTAAATATATCT---CAAT 3448
Db 1193 euHisGlnProLeuSerGluAlaGlnArgLysSerLysSerLeuLysIleAsnLeuAsnT 1213
QY 3449 TAGCAGGAGATGCTCTAAGAGA---AATACATTTGTTTCAGAACATCAACAGAGA---CA 3502
Db 1213 yrAlaGlyAsp-AlaLeuArgGluAsnThrLeuValSerGluHis-AlaGlnArgAspGl 1232
QY 3503 ACCTGAAAACAGTGTCAAT---GAAGGAGGAGACACATGATCAAAAC-----GACAA 3553
Db 1232 rArgGluThrGlnCysGlnMetLysGluAlaGluHisMetTyrGlnAsnGluGlnAspAs 1252
QY 3554 GATAATGGAACAAACACCTGACAGCAGAGTCTCTAGTCAGAAATTTTCAACTACAAG 3613
Db 1252 nValAsnLysHisThrGluGlnGlnGluSerLeu-AspGlnLysLeuPheGlnLeuGlns 1272
QY 3614 CAAAAATTTGGCTTCACAGCAAT---AGTCATGCACATAGAAAGCTGCACACAAAG 3669
Db 1272 erLysAsnMetTrpLeuGlnGlnLeuValHisAlaHisLys-----LysA 1288
QY 3670 CAGATAACAAAT-----GATATTCTATTTCTTTC---AGAGGAAATGCAACAC 3711
```

```
Db 1288 laAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLysMetGlnHisH 1308
QY 3712 ATCTCTCTAAAGAGAAAAAGAGAGATATT---AATTACAATACCATTTAAAAACCGTAT 3768
Db 1308 isLeuLeuLysGluLysAsnGlu-GluIlePheAsnTyrAsnAsnHisLeuLysAsnArg 1327
QY 3769 ATT---CAATATGCAAAAAGAGAAA 3789
Db 1328 IleTyrGlnTyrGluLysGluLys 1335
RESULT 5
US-10-124-805-565
; Sequence 565, Application US/10124805
; Publication No. US20030166022A1
; GENERAL INFORMATION:
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C12
; CURRENT APPLICATION NUMBER: US/10/124,805
; CURRENT FILING DATE: 2002-04-15
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 565
; LENGTH: 1341
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-124-805-565
Alignment Scores:
Pred. No.: 8,898-110 Length: 1341
Score: 1234.00 Matches: 679
Percent Similarity: 59.18% Conservative: 146
Best Local Similarity: 48.71% Mismatches: 376
Query Match: 17.41% Indels: 207
DB: 14 Gaps: 115
US-09-602-362E-22 (1-4115) x US-10-124-805-565 (1-1341)
QY 96 CAAAGGAAAGAGACATCACTTAATATACAGAG---CCAGAGAGAGACTGCTCTAAC 152
Db 3 LysArgLysLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu-- 21
QY 153 TGGCGCTGTGTCATCGCCT-----CAGGAAGTAGTAACATTTC---GGTA 194
Db 22 -----HisTrpAlaCysValAsnGlyHisGluGluValValThr-PheLeuValA 38
QY 195 GACAGAGAGTCCAGCTGACGTCTTTC---TGCGAACACGGACACCTCTGTATGAGGC-T 250
Db 38 spArgLysCysGlnLeuAspValLeuAspValLeuAspGlyGluHisArgThrProLeuMetLysAlaL 58
QY 251 TACATGCCACAGGAGGCTTTCG---AAATATTTCATAGATTTC-GGTCCGATATAAATC 306
Db 58 euGlnCysHisGlnGluAlaCysAlaAsnIleLeuIleAspSerGlyAlaAspIleAsnL 78
QY 307 TCGTG---ATGTGTATGCAACATGCTTCCATTATTCGTTTATAGTGATTTTGTGAGT- 362
Db 78 euValAspValTyrGlyAsnMetAlaLeuHisTyrAlaValTyrSerGluIleLeuSerV 98
QY 363 ---GGTGGCAAACTGTGTC-CATGGTCAGATTCAGATGC---CAACAAGGCTCCCTCA 416
Db 98 aIvalAlaLysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeuT 118
QY 417 CACCACCTTTTACTACCAT---AACGAAAGAAAGTAGCAATTGTGGAATTTT---CCTGAT 470
Db 118 hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuLeuIle 137
QY 471 AAAAATGCAAAATGCAATGCGTTA---ATAAGTTAAATGCAACACCTCATGCTGTGTAT 527
Db 138 LysAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAlaVal 157
```

[illegible]

| | | | | | | | | | | | |
|----|------|--------------|-------------|---------------|----------------|-----------------|-----------------|-----------------|---------------|------------|------|
| QY | 1499 | TCTTGGGATTC | TAGAGTCTCTG | --- | GAGACTGTTT | CAGAAG | --- | ATTGTGTTT | ACC | 1555 | |
| DB | 509 | SerTrpAspSer | -GluSerLeu | CysGluThrVal | SerGlnLys | AspVal | CysLeu | ProLys | 528 | | |
| QY | 1553 | GGCTACACAT | CAAAAGAAAT | TAGATAAAAA | --- | AATGGAAAA | --- | TAGACAGTCC | 1606 | | |
| DB | 528 | salathHisGln | LysGlu | -IleAsp | LysIle | AsnGlyLys | LeuGluGlu | SerProAsn | 548 | | |
| QY | 1607 | AAAGTGTCTT | CTGAAGCT | ACTTCGGAAAT | --- | GAAATTTCT | TATTCAC | TAAAGCC | 1660 | | |
| DB | 548 | ysAspGlyLeu | LeuLysAla | ThrCys | -GlyMet | LysValSer | IleProThrLys | AlaLeu | 567 | | |
| QY | 1661 | AGATTGAAC | ATGC | CAAACTTT | CAAGCGAGCTCC | --- | GGGAAGCC | ACTTGCT | 1714 | | |
| DB | 568 | GluLeuLys | AspMetGln | ThrPheLys | AlaGlu | -ProPro | GlyLysPro | SerAlaPhe | 587 | | |
| QY | 1715 | GAGCTCCACT | GAAATCAAA | AGTCTGTCC | CAATAAAGCCTTGG | ----- | AATGAAAA | 1765 | | | |
| DB | 587 | uProAla | ThrGlu | MetGln | LysSerValPro | AsnLys | AlaLeuGlu | LysAsnGlu | 606 | | |
| QY | 1766 | ATGAAAACAT | GAGGAGAT | GTAGAA | --- | CTCCCATC | AGATCCAA | CAAGACTATA | 1822 | | |
| DB | 607 | --- | GlnThrTrp | ArgAla | Asp | -GluLeu | LeuProSerGlu | SerLysGlnLys | 625 | | |
| QY | 1823 | GA AAA | --- | TTCTGG | CACTCGAGTCTCTG | --- | TAGACTGTTT | CACAGAAGGATG | 1875 | | |
| DB | 625 | luGlu | AsnSerTrp | AspThr | -GluSer | LeuCysGlu | ThrValSerGln | LysAspValCys | 644 | | |
| QY | 1876 | TTACAAG | --- | CTGCCT | CAAAAAGATAG | TAAATAAT | TGGAAAT | TAGAGGGT | 1929 | | |
| DB | 645 | LeuProLys | AlaAla | HisGln | LysGlu | IleAsp | LysIleAsnGly | LysLeuGluGly | 664 | | |
| QY | 1930 | CCCTGTTAA | AGTGCTTCT | GAGCTAACT | --- | GGGAATGAA | AGTTCTATT | TCCAACTAAA | 1986 | | |
| DB | 665 | ProVal | LysAspGly | LeuLeuLys | AlaAsn | CysGly | MetLysValSer | IleProThrLys | 684 | | |
| QY | 1987 | GCCTPAG | --- | ATTGAT | GACATGCAAACT | CTCAAAG | --- | CAGCGCTCC | CGAGCCATCTGC | 2040 | |
| DB | 685 | Ala | -LeuGlu | LeuMet | AspMetGln | ThrPheLys | AlaGlu | ProProGlu | LysProSerAl | 704 | |
| QY | 2041 | CTTCGAGCT | CCATTGAA | --- | TCAAAGTCTGTCC | CAATAAAT | --- | CTTGGAATTA | GA | 2094 | |
| DB | 704 | aPheGlu | ProAla | IleGlu | MetGlnLys | SerValPro | AsnLys | AlaLeuGlu | LysLeuLys | 724 | |
| QY | 2095 | ATGAACAAA | CAATTGAGG | CAGATGAGAA | --- | CTCCCATCA | --- | AATCAAA | CAAGAGACTA | 2148 | |
| DB | 724 | snGluGln | ThrLeu | ArgAla | asp | -GluIle | LeuProSerGlu | SerLysGlnLys | AspTyr | 743 | |
| QY | 2149 | TGAGAAAGT | TTCTTGG | ATCTGAGTCTCTG | --- | TAGACTGTTT | CCAGAAGGATGGT | --- | 2203 | | |
| DB | 744 | GluGlu | SerSerTrp | AspSer | -GluSer | LeuCysGlu | ThrValSerGln | LysAspValCys | 763 | | |
| QY | 2204 | TTTACC | CAGGCTAC | ATCAAAAGAAAT | TAGATAAAAA | --- | TAATGG | AAATAGAGAAGT | 2259 | | |
| DB | 763 | sLeuProLys | AlaThrHis | GlnLysGlu | IleAsp | LysIleAsnGln | LysLysLeuGluGlu | 583 | | | |
| QY | 2260 | TCT | -GATAAT | GAGGTTTCT | GAAGCTCCCT | --- | GAGAAATG | AAAGTTCTATT | CCACTAAA | 2315 | |
| DB | 783 | xProAsp | AsnAspGly | PheLeuLys | AlaProCys | ArgMet | LysValSerIle | ProThrLys | 803 | | |
| QY | 2316 | GCCT | --- | TGAATT | GATGGCATG | CAAACTTT | CAAGCAGCGCTCCCG | --- | AGAGCCATCTG | 2369 | |
| DB | 803 | sAlaLeu | GluLeuMet | AspMetGln | ThrPheLys | AlaGlu | ProProGlu | LysProSerAl | 823 | | |
| QY | 2370 | CTTCGAGC | CTCAVTTG | AAAT | --- | GAAAAGTCTGT | TCCAAATAAA | --- | CTTGGAAATTAAG | 2423 | |
| DB | 823 | aPheGlu | ProAla | Ile | -GluMet | GlnLysSerValPro | AsnLys | AlaLeuGlu | Lys | 842 | |
| QY | 2424 | AATGA | CAACAT | TGAGAC | AGATACAT | ----- | TTCCCTT | CAGNATCA | AAACAAGAGG | 2477 | |
| DB | 843 | AsnGlu | GlnThr | --- | -LeuArg | AlaAspGln | MetPhePro | SerGluSerLysGln | Lys | 861 | |
| QY | 2478 | TTGAGAAA | --- | TTCTG | GGATCTT | GAAGTCT | --- | CCGTGG | ACTGTTTCA | CAGAGGATGG | 2531 |

Db 862 ValGluGluAsnSerTrpAspSerGlu-SerLeuArgGluThrValSerGlnLysAsp-V 881
QY 2532 TGTGTACC---CAGGCTACACACAAAAAGAGGATAAAAT---AGTGGAATAATTAGAAAG 2585
Db 881 aLysValProLysAlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluA 901
QY 2586 ATTCACTAGCCTATCAAAAT---CTTGTCAGTTCAT---CTTGTAAGACGAGGAA 2639
Db 901 spSerThrLeuSerLysIleLeuAspThrValHisSerCysGlnArgAlaArgGluL 921
QY 2640 CTTCAAAAAGACACTGTGAACA-CGTACAGGAAAATGCAACAAT---GAAAAAGAGTTT 2695
Db 921 euGlnLysAspHisCysGlnArgThrGlyLysMetGlnMetLysLysLysPheC 941
QY 2696 GTGTACTGAAAAGAAC---TCTCAGAACAAAGAAATAAATCACAGTAGAGAA---CCA 2749
Db 941 ysValLeuLysLysLysLeuSerGluAlaLysGluLysSerGln-LeuGluAsnGln 960
QY 2750 AAAGTTAAATGGACACAGAG---TCTCAGGTAGATTGACTTAAACCAA---GAAAAG 2803
Db 961 LysValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGlu 980
QY 2804 AGAAGAGAGAAATGCCGATATATTAAAGAAAAA---TTAGGAAGAATTAG 2851
Db 981 LysArg-ArgAsnAlaAspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIle-G 1000
QY 2852 AAGAATCGAGACGACATAGAAAGTTAGAGTGAACACACACTTGAAG---GCTCTC 2908
Db 1000 LuGluGlnHisArg-----LysGluLeuGluValLysGlnGlnLeu-GluGlnAlaLeu 1017
QY 2909 AGATCAACAG---ATAAGAATTCGAAGCTGTAGAAAGTAATTTCAATAGCTTTCTCA---AC 2962
Db 1018 ArgIleGlnAspIleGluLysSerValGluSerAsnLeuAsnGln-ValSerHisTh 1037
QY 2963 TCATGAATGAAA---ATTATCTTACATGAATAATTCATGCTGTGTAAGAAAGAAATGCCA- 3017
Db 1037 rHisGluAsnGluAsnTyrLeuLeuHisGluAsnCysMetLeuLysLysGluIleAlaMe 1057
QY 3018 -TGCAAACTGGAATAGCCACACTA---AACACCAATCCAGAAAGGAAAAATAATC-T 3072
Db 1057 tLeuLysLeuGluIleAlaThrLeuLysHis-GlnTyrGlnGluLysGluAsnLysTyrP 1077
QY 3073 TTGAGACTTAAGATTTAAAGA---AAAGATGCTGAACCTTAG-ATGACCTTAAACTGA 3128
Db 1077 heGluAspIleLysIleLeuLysGluLysAsnAlaLeuGlnMetThrLeuLysLeuL 1097
QY 3129 AAAGGAATCATTAATAAG---GGATCTCAATATGTGGCAGCT---AAAGTTCTGATA 3182
Db 1097 ysGluGluSerLeuThrLysArgAlaSerGlnTyrSerGly-GlnLeuLysValLeulle 1116
QY 3183 GCTGAGAAACAAATGCTCATCTTAAATTGAGGAAACACAGACAAAGAAAA---CTAGA 3236
Db 1117 AlaGlu-AsnThrMetLeuThrSerLysLeuLysGluLysGlnAspLysGluIleLeuGl 1136
QY 3237 GCAGAAATTTGAATA-CACCATCTGACTGCTTCGCTGTACAGC-----CA 3283
Db 1136 uAlaGluIleGluSerHisHisPro-----ArgLeuAlaSerAlaValGlnAspHi 1153
QY 3284 TGATCAATTTGACATCATCAAGAAAAAGTAAGA---ACCTGCTTCCACATTTAG---GAGAT 3337
Db 1153 shspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl 1173
QY 3338 GCTGTTTCAAGAAAAATGATGTTGATGTG---AGAGTACGATATTAACATGAGTGCT 3394
Db 1173 aCysLeuGlnArgLysMetAsnValAspValSerSerThrIleTyr-AsnAsnGluValL 1193
QY 3395 CCATCACCCTTTC---TCAGCTCAAGGAAATCCAAACCTAAATAATTTACT---CAAT 3448
Db 1193 euHisGlnProLeuSerGluAlaGlnArgLysSerLysSerLeuLysIleAsnLeuAsnT 1213
QY 3449 TAGCAGGAGATGCTTAAGAGA---AATCATTTGTTTCAGAACATCACAAGAGA---CA 3502

Db 1213 yrAlaGlyAsp-AlaLeuArgGluAsnThrLeuValSerGluHis-AlaGlnArgAspGl 1232
QY 3503 ACTGTAAAAACAGTGTCAAT---GAAGGAACGCAACACATGTATCAAAAC-----GACAA 3553
Db 1232 nArgGluThrGlnCysGlnMetLysGluAlaGluHisMetTyrGlnAsnGluGlnAspAs 1252
QY 3554 GATAATGGACAAACACCTGAACAGCAGAGTCTCTAGTCAGAAATTTATTTCAACTACAAG 3613
Db 1252 nValAsnLysHisThrGluGlnGlnSerLeu-AspGlnLysLeuPheGlnLeuGlns 1272
QY 3614 CAAAAATTTGGTTCACAGCAATT---AGTCATGCACATAGAAAGCTGCAACAAAAAG 3669
Db 1272 erLysAsnMetTrpLeuGlnGlnLeuValHisAlaHisLys-----LysA 1288
QY 3670 CAGATTAACAAAT-----GATATTCATTTCTTG---AGAGGAAATGCAACAC 3711
Db 1288 laAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLysMetGlnHisH 1308
QY 3712 ATCTCTAAAGAGAAAAAGAGAGATATT---AATTCAATACCATTTTAAAAAACCCTAT 3768
Db 1308 isLeuLeuLysGluLysAsnGlu-GluIlePheAsnTyrAsnAsnHisLeuLysAsnArg 1327
QY 3769 ATT---CAATATGAAAAAGAGAAA 3789
Db 1328 IleTyrGlnTyrGluLysGluLys 1335
RESULT 6
US-10-007-805-573
; Sequence 573, Application US/10007805
; Publication No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 573
; LENGTH: 1349
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-007-805-573
Alignment Scores:
Pred. No.: 1,03e-108 Length: 1349
Score: 1223.00 Matches: 677
Percent Similarity: 59.07% Conservative: 150
Best Local Similarity: 48.36% Mismatches: 380
Query Match: 17.26% Indels: 207
DB: 13 Gaps: 115
US-09-602-362E-22 (1-4115) x US-10-007-805-573 (1-1349)
QY 96 CAAAGGAGAGACACATCAACCTTAATATACAGAG---CCGAGAGAGACTGCTCTAAC 152
Db 11 LysArgLysLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu-- 29
QY 153 TGGGCCTGCTCAATGSCCT-----GAGGAAGTAGTAACATTTTC---GGTA 194
Db 30 -----HisTrpAlaCysValAsnGlyHisGluGluValValThr-PheLeuValA 46

QY 195 GACAGAAGTGCAGTCAGCTCTTGG---TGGCAACACGACACCTCTGATGAAGGC-T 250
Db : : : : :
46 spArgLysCysGlnProaspValLeuaspGlyGluHisArgThrProLeuMetLysAlaL 66
QY 251 TACATGCCACAGAGGCTTGGC---AAATATTTGATAGATTC-GGTGCCGATATAATC 306
Db : : : : :
66 euGlnCysHisGlnGluAlaCysAlaasnIleLeuIleAspSerGlyAlaAspIleasnL 86
QY 307 TCGTG---ATGTGTATGCAACATGGCTTCCATTATGGTTTATAGTGGATTTTCAGT- 362
Db : : : : :
86 euValaspValTyGlyAsnMetAlaLeuHisTyAlaValTySerGluIleLeuSerV 106
QY 363 ---GGTGGCAACTGCTGTC-CATGGTCAGATCGAAGTGC---CAACAAGGCTGCCTCA 416
Db : : : : :
106 alValAlaLysLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeuT 126
QY 417 CACCACTTTTACTACCAT---AAGCAAGAAGTGCAGCAATGTGCAATTTT---GCTGAT 470
Db : : : : :
126 hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuLeuLe 145
QY 471 AAAAATGCAATGCAATGCTTA---ATAAGTTTAAATGCACCCCTCATGCTGCTGAT 527
Db : : : : :
146 LysAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAlaVa 165
QY 528 G---TCTGGATCATCAGATAGTGTGATGCTTCTTAGCAAAA---TGTCAGCGTCTTT-- 579
Db : : : : :
165 lCysHisGlyLeuSerGluIleValGlyMetLeuLeuGlnGlnAsnValaspValPheAl 185
QY 580 -GTGCAGATATAGTGGAGTAAGTGCAGAACAT---ATGCTGTTACTGTGATTTCTCACA 635
Db : : : : :
185 aAlaAspIle-CysGlyValThrAlaGluHisTyAlaValThrCysGlyPheHisI 205
QY 636 TTCATACAAA---TTATGAATATATACGAAAATTCAAAATCA---TCAATATACAA 689
Db : : : : :
205 leHisGluGlnIleMet-GlutyrIleArgLysLeuSer-LysAsnHisGlnAsnThrAs 224
QY 690 T---CAGAGAAGCACTGTCAGGAAACCTGATGAGCTGCACCCCT---GGCGGAAGACAC 743
Db : : : : :
224 nProGluGlyThrSerAlaGlyThrProaspGluAlaAla-ProLeuAlaGluArgThrP 244
QY 744 CTGACAGCTGAAGACT---GGTGGAAAAACCTGATGAGCTGCACCCCT---GGTGGGA 797
Db : : : : :
244 roAspThrAlaGluSerLeuValGluLysThrProaspGluAlaAla-ProLeuValGlu 263
QY 798 AAGACACCTGACAGCTGAAAGCT---GGTGGAAAAACCTGATGAGCTGCATCT--- 851
Db : : : : :
264 ArgThrProaspThrAlaGluSerLeuValGluLysThrProaspGluAlaAlaSerLeu 283
QY 852 TGGTGGAGGAACATCTGCACAAATTCATGTTGGAGAAGG---ACATCTCGAAAGTTTGA 908
Db : : : : :
284 Val-GluGlyThrSerAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheGl 303
QY 909 ACGTCAGC---AGAAAACACCTAGGAATTCAGATCC---TGCAAAAAACATCTGAGA 962
Db : : : : :
303 uGln-SerAlaGluGluThrProArgGluIleThrSerProAlaLysGluThrSerGluL 323
QY 963 AATTTCAG---GGCCAGCAAGAGACCTAGAGATCGCAGGAGAGAA---AAAGAAGAC 1016
Db : : : : :
323 ysPheThrTrpPro-AlaLysGlyArgProArgLysIleAlaTrpGluLysLysGluasp 342
QY 1017 ACACCTAGGGAATATAGTCCGAAAGAAAGAACTAGAGAA-----TTA 1061
Db : : : : :
343 ThrProArgGluIleMetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAla 362
QY 1062 GGTGGCAGAAAAGGAAGACCTAGGAAGCCATGGGAGAAAAGAAA---CCCTGTAAA 1118
Db : : : : :
363 LysGlyArgProArgLysIle-----AlaTrpGluLysLysGluThrProValLys 379
QY 1119 GATGATGCGT-GCAAGAGTAACTATTAATAAATAAG---TTTGGAAAAAGAAATCT 1174
Db : : : : :
380 ThrGlyCysValAlaArgValThrSerAsnLysThrLysValLeuGluLysGlyArgSer 399
QY 1175 AAGATGTT-GCATGTCCACAAAG---AATCATCTACAGCAAGTCCCATGATCAGAG 1230

Db 400 LysMetIleAlaCysProThrLysGluSerSerThrLysAlaSerAlaAsnAspGlnArg 419
QY 1231 ---TCCCATCAGATCCAAACAAGGAAGATGAGATATTTCTTT-GATTCTCGGTCTCT 1286
Db : : : : :
420 PheProSerGluSerLysGlnGluaspGluGluTrpSerCysAspSerArg---Ser 438
QY 1287 TTG-----AAGTTCTGCAAGATTCAAGTG---GTATACCTGAGTCTATATATAAAAGT 1337
Db : : : : :
439 LeuPheGluSerSerAlaLysIleGlnValCysIleProGluSerIleTyGlnLysVal 458
QY 1338 AATGAGATAAATAAGAG---TAGAAGGCTCTCTAAAGCCATCTGCTTCAAGCGCCA 1394
Db : : : : :
459 MetGluIleAsnArgGluValGluGluProProLysLysProSerAlaPheLysProAla 478
QY 1395 TTGAAGAGCAAAA---CTCTTCCAAATAAGCCTTTG-----AATCAAGAATGAACA 1442
Db : : : : :
479 IleGluMetGlnAsnSerValProasnLysAlaPheGluLeuLysasnGluGln----- 496
QY 1443 ACATTGAAGCAGATCCGTG-TTCCACCG---AATCCAAAACAAGGACTATAAGAAAT 1498
Db : : : : :
497 ThrLeuArgAlaAspProMetPheProGluSerLysGlnLysAspTyGluGluasn 516
QY 1499 TCTTGGATTCTAGATCTCTG---GAGACTGTTTACAGAAGG---ATTGTGTTTACCAA 1552
Db : : : : :
517 SerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLy 536
QY 1553 GGCTACACATCAAAAGAATAGATAAAA---AATGAAAA---TAGAAGATTCCTAAT 1606
Db : : : : :
536 salaThrHisGlnLysGlu-IleAspLysIleasnGlyLysLeuGluGluSerProasnL 556
QY 1607 AAGTGGTCTTCTGAGGCTACTCGAAT---GAAATTTCTATTCCTCCACTAAGCCT--- 1660
Db : : : : :
556 ysAspGlyLeuLeuLysAlaThrCys-GlyMetLysValSerIleProThrLysAlaLeu 575
QY 1661 AGAATTGAAGACATGCAAACTTTCAAGCGAGCCCTCC---GGGAAGCCATCTGCT---TC 1714
Db : : : : :
576 GluLeuLysaspMetGlnThrPheLysAlaGlu-ProProGlyLysProSerAlaPheGl 595
QY 1715 GAGCTTCCACTGAAATCAAAAGTCTGTCCCAATAGCCCTCG-----AATGAAA 1765
Db : : : : :
595 uProAlaThrGluMetGlnLysSerValProasnLysAlaLeuGluLeuLysasnGlu- 614
QY 1766 ATGAAAAACATGGAGGCAGATGAGAA---CTCCCATCAGAAATCCAAACAAGGACTATAA 1822
Db : : : : :
615 ---GlnThrTrpArgAlaasp-GluIleLeuProSerGluSerLysGlnLysAspTyG 633
QY 1823 GAAAA---TTCTGGGATACTGGAGTCTCG---TAGACTGTTTTCACAGAAGATTG-TGT 1875
Db : : : : :
633 luGluasnSerTrpaspThr-GluSerLeuCysGluThrValSerGlnLysaspValCys 652
QY 1876 TTACCAAGG---CTCGCTCAAAAAGATAGATAAAATATCGAAAATTTAGAAGGT--- 1929
Db : : : : :
653 LeuProLysAlaAlaHisGlnLysGluIleAspLysIleasnGlyLysLeuGluGlySer 672
QY 1930 CCTGTGTAAGTGGTCTTCTGAGGCTAAT---GGGAATGAAAAGTTTCTATTCCAACTAAA 1986
Db : : : : :
673 ProValLysaspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLys 692
QY 1987 GCCTAGA---ATTGATGACATGCAACTTCAAG---CAGGCTCCCGAGAGCCATCTGC 2040
Db : : : : :
693 Ala-LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl 712
QY 2041 CTTGAGCCTCCATTGAAA---TCAAAAGTCTGTCCAAATAAAC---CTTGAATTAAGA 2094
Db : : : : :
712 aPheGluProAlaIleGluMetGlnLysSerValProasn-LysAlaLeuGluLeuLysA 732
QY 2095 ATGAACAACATTGAGGCAGATGAGAA---CTCCCATCA---AATCCAAACAAGGACTA 2148
Db : : : : :
732 snGluGlnThrLeuArgAlaasp-GluIleLeuProSerGluSerLysGlnLysaspTy 751
QY 2149 TGAGAAAGTCTTGGGATTCTGAGTCTCTG---TAGACTGTTTCCAGAAAGATCGTG--- 2203

Db 752 GluGluSerSerTrpAspSer-GluSerLeuGluThrValSerGlnLysAspValCy 771
 QY 2204 -TTTACCAGGCTACACATCAAAAGATAGATAAAA---TAATGGAATAGAGAGTC 2259
 Db 771 sLeuProLysAlaThrHisGlnLysGluLeuAspLysLeuGluGluSe 791
 QY 2260 TCT-GATTAATGAGGTTTCTGAGGCTCCCT---GAGNATGAAAGTTCTATTCCACTAAA 2315
 Db 791 rProAspAsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLy 811
 QY 2316 GCCT---TCGATTGATGCATGCAAACTTCAAGAGCGCTCCCG---AGAGCCATCTG 2369
 Db 811 sAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAl 831
 QY 2370 CTTGAGGCTGATTGAAAT---GAAAGTCTGTTCCAAATAAAC---CTTGGAAATTAAG 2423
 Db 831 aPheGluProAlaIle-GluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLys 850
 QY 2424 AATGACACACATTGACAGATCAGAT---TTCCCTTCAGAAATCAAAACAAAGAGG 2477
 Db 851 AsnGluGlnThr---LeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLys 869
 QY 2478 TTGAGAAA---TTCTGGGATTCGAAGTCT---CCGTGGACTGTTTCACAGAGGATGG 2531
 Db 870 ValGluGluAsnSerTrpAspSerGlu-SerLeuArgGluThrValSerGlnLysAsp-V 889
 QY 2532 TGTGTACC---CAGCTACACAAAAGAAAGAGGATAAAAT---AGTGGAAATTAAG 2585
 Db 889 aLysValProLysAlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGlu 909
 QY 2586 ATTCACTAGCTATCAAAAT---CTTGTACAGTTTCAT---CTTGTGAAAGGCAAGGAA 2639
 Db 909 spSerThrSerLeuSerLysIleLeuAspThrValHisSerCysGluArgAlaArgGlu 929
 QY 2640 CTTCAAAAGACACTGTGAACA-GTACAGAAAATGGAACAAT---GAAAGAGGTTTT 2695
 Db 929 euGlnLysAspHisCysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheC 949
 QY 2696 GTGTACTGAAAGAAC---TGTCAAGAAAAGAAATAAATCACAGTAGAGAA---CCA 2749
 Db 949 ysValLeuLysLysLysLeuSerGluAlaLysGluLeuLysSerGln-LeuGluAsnGln 968
 QY 2750 AAAGTTAAATGGACAAAGAG---TCTCAGTGTAGATGCTTAAACCAA---GAAAG 2803
 Db 969 LysValLysTrpGluGlnGluLysCysSerValArgLeuThrLeuAsnGlnGluGlu 988
 QY 2804 AGAGAGAAATGCCGATATATTAAGAAAAA---TTAGGAAGAAATAG 2851
 Db 989 LysArg-ArgAsnAlaAspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIle-G 1008
 QY 2852 AAGATCGAGAGCAGCATAGAAAGTTAGAAAGTGAACACACAACTTGAAG---GCTCTC 2908
 Db 1008 luGluGlnHisArg---LysGluLeuGluValLysGlnGlnLeu-GluGlnAlaLeu 1025
 QY 2909 AGATACAG---ATPAGAAATGAGCTGTAGAAAGTANTTGAATAGTTCTCA---AC 2962
 Db 1026 ArgIleGlnAspIleGluLeuLysSerValGluSerAsnLeuAsnGln-ValSerHisTh 1045
 QY 2963 TCATCAATGAAA---ATTATTCTTACATGAAATGCTGTTGAAAAGGAATGCCA--- 3017
 Db 1045 rHisGluAsnGluAsnTyLeuLeuHisGluAsnCysMetLeuLysLysGluIleAlaMe 1065
 QY 3018 -TGCAAACTGAATAGCACACTA---AACACCAATCCAGAAAAGAAAATAATC-T 3072
 Db 1065 tLeuLysLeuGluIleAlaThrLeuLysHis-GlnTyGlnGluGlyGluAsnLysTyP 1085
 QY 3073 TTGAGACTTAAGATTTAAAGA---AAAGATGCTGAACTTAG-ATGACCCCTAAACTGA 3128
 Db 1085 heGluAspIleLysLysLysGluLysAsnAlaGluLeuGlnMetThrLysLysLeu 1105
 QY 3129 AAAGGAATCACTACTAAAG---GGATCTCAATATGTGGGAGCT---AAGTCTGATA 3182
 Db 1105 ysGluGluSerLeuThrLysArgAlaSerGlnTySerGly-GlnLeuLysValLeuIle 1124

QY 3183 GCTGAGAAACAATGCTCATCTTAATTTAGGAAAAACAGACAAAGAAA---CTAGA 3236
 Db 1125 AlaGlu-AsnThrMetLeuThrSerLysLeuLysGlnAspLysGluIleLeuGl 1144
 QY 3237 GCAGAAATTTGAATA-CACCATCTGAGCTGGCTTCGCTGTACAAGC---CA 3283
 Db 1144 uAlaGluIleGluSerHisHisPro-----ArgLeuAlaSerAlaValGlnAspHi 1161
 QY 3284 TGATCAATTTGATGACATCAAGAAAAAGTAAGA---ACCTGCTTCCACATTTAG---GAGAT 3337
 Db 1161 sAspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl 1181
 QY 3338 GCTGTTTGCAGAAAGAAATGAATGTTGATGTG---AGAGTACGATATTAAACATAGTGCT 3394
 Db 1181 aCysLeuGlnArgLysMetAsnValAspValSerSerThrIleTy-AsnAsnGluValL 1201
 QY 3395 CCATCACCACCTTTC---TGAGCTCAAGGAAATCCAAAACCTTAAATTTATCT---CAAT 3448
 Db 1201 euHisGlnProLeuSerGluAlaGlnArgLysSerLysSerLeuLysIleAsnLeuAsnT 1221
 QY 3449 TAGCAGGAGATGCTTAAGAGA---AATACATTTGTTTCAGACATCATCAAAAGAGA---CA 3502
 Db 1221 yrAlaGlyAsp-AlaLeuArgGluAsnThrLeuValSerGluHis-AlaGlnArgAspGl 1240
 QY 3503 ACCTGAAAAACAGTGTCAAT---GAAGGAAGCGAAACACATGTATCAAAAC---GACAA 3553
 Db 1240 nArgGluThrGlnCysGlnMetLysGluAlaGluHisMetTyGlnAsnGluGlnAspAs 1260
 QY 3554 GATAATGGAACAAACACCTGAAACAGACAGAGTCTTAGTCAGAAATTTTCAACTACAAG 3613
 Db 1260 nValAsnLysHisThrGluGlnGlnGlnSerLeu-AspGlnLysLeuPheGlnLeuGlnS 1280
 QY 3614 CAAATATTGCTTGCCTCACAGCAATT---AGTCATGCGCATAGAAAGCTGCAACAAAAAG 3669
 Db 1280 erLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaHisLys-----LysA 1296
 QY 3670 CAGATAACAAT-----GATATTCATTCTTCTG---AGAGGAATGCAACAC 3711
 Db 1296 laAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLysMetGlnHisH 1316
 QY 3712 ATCTCTTAAAGAGAAAAGAGAGAGATATT---AATTACAATACCATTTAAAAACCGTAT 3768
 Db 1316 isLeuLeuLysGlnLysAsnGlu-GluIlePheAsnTyAsnAsnHisLeuLysAsnArg 1335
 QY 3769 ATT---CAATATGAAAGAGAAAGAGAAACAGAAACTCATG 3807
 Db 1336 ileTyGlnTyGluLysGluLysAlaGluThrGluValIle 1349

RESULT 7
 US-10-076-622-573
 ; Sequence 573, Application US/10076622
 ; Publication No. US20030023036A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Houghton, Raymond L.
 ; APPLICANT: Sleath, Paul R.
 ; APPLICANT: Persing, David H.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; FILE REFERENCE: 210121.470C11
 ; CURRENT APPLICATION NUMBER: US/10/076,622
 ; CURRENT FILING DATE: 2002-02-13
 ; NUMBER OF SEQ ID NOS: 627
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 573
 ; LENGTH: 1349
 ; TYPE: PRN
 ; ORGANISM: Homo sapiens
 US-10-076-622-573
 Alignment Scores: 1,03e-108 Length: 1349
 Pct. No.: 1223.00 Matches: 677
 Score:

| | | | |
|---|--------|--|------|
| Percent Similarity: | 59.07% | Conservative: | 150 |
| Best Local Similarity: | 48.36% | Mismatches: | 380 |
| Query Match: | 17.26% | Indels: | 207 |
| DB: | 14 | Gaps: | 115 |
| US-09-602-362E-22 (1-4115) x US-10-076-622-573 (1-1349) | | | |
| QY | 96 | CAAGGAGAGAGACATCACTTAATATACAGAG--CCACAGAGAGACTGCTCTAAC | 152 |
| DB | 11 | LysArgLysLysThrIleasnLeuasnIleGlnaspAlaGlnLysArg-ThrAlaLeu-- | 29 |
| QY | 153 | TGGGCGCTGCTCAATGGCCT-----GAGGAAGTAGTAACATTTCTC--GGTA | 194 |
| DB | 30 | -----HisTrpAlaCysValasnGlyHisGluGluValValThr-PheLeuValA | 46 |
| QY | 195 | GACAGAAGTCCAGCTGAGCTCTG---TGCGACACAGCAGCCTCTGATGAAGC-T | 250 |
| DB | 46 | sPArgLysCysGlnProaspValLeuaspGlyGluHisArgThrProLeuMetLysAlaL | 66 |
| QY | 251 | TACAATGCCACAGGAGGCTTTGC-----AAATATTGATAGATTCTC-GGTCCCGATATAATC | 306 |
| DB | 66 | eugInCysHisGlnIleAlaCysAlaAsnIleLeuIleaspSerGlyAlaaspIleAsnL | 86 |
| QY | 307 | TCGTG---ATGTGATGCAACATGCTTCCATTATGCGTTTATAGTGAGTTTGTCTAGT- | 362 |
| DB | 86 | euValaspValTyroGlyAsnMetAlaLeuHisTyAlaValTySerGluIleLeuSerV | 106 |
| QY | 363 | -GGTGGCAAACTGCTGTC-CATGTGTCAGTATCGAAGTGC---CAACAAGGCTGCCTCA | 416 |
| DB | 106 | alValAlaLysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeuT | 126 |
| QY | 417 | CACCACTTTTACTACCAT---AACGAAGAAGTGAAGCAATGTGGAATTTT---GCTGAT | 470 |
| DB | 126 | hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuIle | 145 |
| QY | 471 | AAAAATGCAATGCAATGCGTTA---ATAAGTTAAATGCACACCTCATGCTCTGTAT | 527 |
| DB | 146 | LysasnAlaAsnAlaAsnAlaValasnLys-TyrlsCysThrAlaLeuMetLeuAlaVa | 165 |
| QY | 528 | G---TCTGGATCATCAGAGATAGTGTGATGCTTCTTACCAAA---TGTGACGCTCTTT-- | 579 |
| DB | 165 | lCysHisGlyLeuSerGluIleValGlyMetLeuLeuGlnGlnAsnValaspValPheAl | 185 |
| QY | 580 | -GTCAGATATAGTGAGTACTCAGACAT---ATGCTGTACTGTGATTTCTCACA | 635 |
| DB | 185 | alaaspIle-CysGlyValThrAlaGluHisTyAlaValThrCysGlyPheHisIst | 205 |
| QY | 636 | TTCTATAACAAA---TTATGAATATATACGAAATATCAAAATCA---TCAAAATACCA | 699 |
| DB | 205 | leHisGluGlnIleMet-GluTyrlleargLysLeuSer-LysasnHisGlnasnThrAs | 224 |
| QY | 690 | T---CAGNAGGAACCTGCAGAAACCTGATGAGCTGCACCTT---GGCGGAAGACAC | 743 |
| DB | 224 | nProGluGlyThrSerAlaGlyThrProaspGluAlaAla-ProLeuAlaGluArgThrP | 244 |
| QY | 744 | CTGACAGCTGAAAGCT---GGTGGAAAAACACCTGATGAGCTGCACCTT---GGTGA | 797 |
| DB | 244 | roaspThrAlaGluSerLeuValglulysThrProaspGluAlaAla-ProLeuValGlu | 263 |
| QY | 798 | AAGACACTGACGGCTGAAGCT---GGTGGAAAAACACCTGATGAGCTGCATCT---851 | |
| DB | 264 | ArgThrProaspThrAlaGluSerLeuValglulysThrProaspGluAlaAlaSerLeu | 283 |
| QY | 852 | TGGTGGAGGACACTCAGCAATTCATGTTGGAGAAGG---ACATCTGGAAGTTGGA | 908 |
| DB | 284 | Val-GluGlyThrSeraspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheGl | 303 |
| QY | 909 | ACGTGACG---AGAAAAACACCTAGGAATATAGATCC---TGCAAAAAACATCTGAGA | 962 |
| DB | 303 | uGln-SerAlaGluGluThrProArgGluIleThrSerProAlaLysGluThrSerGluL | 323 |
| QY | 963 | AATTAGG---GGCCACCAAGGAGACCTAGAGATCGCAGGAGAA---AAAGAAGAC | 1016 |
| DB | | | |

| | | | |
|----|------|---|------|
| Db | 323 | ysPheThrTrpPro-AlaLysGlyArgProArgLysIleAlaTrpGluLysLysGluasp | 342 |
| QY | 1017 | ACACCTAGGGAATATGAGTCCCGAAAAAGAAACATGAGAAA-----TTA | 1061 |
| DB | 343 | ThrProArgGluIleMetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAla | 362 |
| QY | 1062 | CGTGGGAGAAAAGAGACCTAGGAAGACGATGGAGAAAAAGAAA-----CCCTGTAAA | 1118 |
| DB | 363 | LysGlyArgProArgLysIle-----AlaTrpGluLysLysGluThrProValLys | 379 |
| QY | 1119 | GATGGATGCTT-GCAAGAGTAACATCTAATAAACTAAAG--TTTGTAAAAAGGAATCT | 1174 |
| DB | 380 | ThrGlyCysValAlaargValThrSerAsnLysThrLysValLeuGluLysGlyArgSer | 399 |
| QY | 1175 | AGATGTT-SCATGTCCACAAAAG---AATCATCTACAAAGCAAGTGCATGATCAGAG | 1230 |
| DB | 400 | LysMetIleAlaCysProThrLysGluSerSerThrLysAlaSerAlaAsnAspGlnArg | 419 |
| QY | 1231 | ---TCCCATCAGATCCAAACAAGGAGATGAAGAATATCTTT-CATTCTCGGCTCTCT | 1286 |
| DB | 420 | PheProSerGluSerLysGlnGluaspGluGluTySerCysaspSerArg---Ser | 438 |
| QY | 1287 | TTG-----AAGTCTGCAAGATTCAAGTG---GTATACCTGAGTCTATATATAAAGT | 1337 |
| DB | 439 | LeuPheGluSerSerAlaLysIleGlnValCysIleProGluSerIleTyroGlnLysVal | 458 |
| QY | 1338 | AATGAGATAAATAAGAG---TAGAAGGCTCTTAAAGCCATCTCCCTTCAAGCCGCA | 1394 |
| DB | 459 | MetGluIleasnArgGluValGluGluProProLysLysProSerAlaPheLysProAla | 478 |
| QY | 1395 | TTGAAAGCAAAA---CTCTTTCCAAATAAGCCCTTTG-----AATGAAAGATGAACA | 1442 |
| DB | 479 | IleGluMetGlnAsnSerValProasnLysAlaPheGluLeuLysasnGluGln----- | 496 |
| QY | 1443 | ACATTGAGCAGATCCGTG-TTCCACCG---AATCCAAACAAGGACTATAAGAAAT | 1498 |
| DB | 497 | ThrLeuArgAlaaspProMetPheProProGluSerLysGlnLysaspTyroGluGluasn | 516 |
| QY | 1499 | TCTTGGGATTCAGAGTCTCTG---GAGACTGTTTACAGAAG---ATTGTGTTTACCAA | 1552 |
| DB | 517 | SerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysaspValCysLeuProly | 536 |
| QY | 1553 | GGCTACACATCAAAAGAATAGATAAAA---AATGGAAAA---TAGAAGAGTCCCTAAT | 1606 |
| DB | 536 | sAlaThrHisGlnLysGlu-IleaspLysIleasnGlyLysLeuGluGluSerProasnL | 556 |
| QY | 1607 | AAAGTGTCTTCTGAGGCTACTCGGAAT---GAAATTTCTATCCACTAAGCCT---1660 | |
| DB | 556 | ysaspGlyLeuLeuLysAlaThrCys-GlyMetLysValSerIleProThrLysAlaLeu | 575 |
| QY | 1661 | AGAATTGAAGACATCAAACTTTCAAGCGACCTCC---GGGAAGCCATCTGCT---TC | 1714 |
| DB | 576 | GluLeuLysaspMetGlnThrPheLysAlaGlu-ProProGlyLysProSerAlaPheGl | 595 |
| QY | 1715 | GAGCTCCACTGAAATCAAAAGTCTGTCCCAATTAAGCCTTG-----AATGAAA | 1765 |
| DB | 595 | uProAlaThrGluMetGlnLysSerValProasnLysAlaLeuGluLeuLysasnGlu--- | 614 |
| QY | 1766 | ATGAAAAACATCGAGCGCAGATGAGAA---CTCCCATCAGAATCCAAACAAGGACTATA | 1822 |
| DB | 615 | ---GlnThrTrpArgAlaasp-GluIleLeuProSerGluSerLysGlnLysaspTyroG | 633 |
| QY | 1823 | GAATA---TTCGGGATGACTGAGTCTCTG---TAGACTGTTTACAGAAGATTG-TGT | 1875 |
| DB | 633 | luGluasnSerTrpaspThr-GluSerLeuCysGluThrValSerGlnLysaspValCys | 652 |
| QY | 1876 | TTACCAAGG---CTCGCTCAAAAGAATAGATAAAATAATGGAATAATAGAGGT---1929 | |
| DB | 653 | LeuProLysAlaAlaHisGlnLysGluIleaspLysIleasnGlyLysLeuGluGlySer | 672 |
| QY | 1930 | CCCTTTAAAGTGTCTTCTGAGGCTAACT---GGGAATGAAAGTCTTATTCCCACTAAA | 1986 |
| DB | 673 | ProValLysaspGlyLeuLeuLysAlaasnCysGlyMetLysValSerIleProThrLys | 692 |

; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C12
; CURRENT APPLICATION NUMBER: US/10/124,805
; CURRENT FILING DATE: 2002-04-15
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 573
; LENGTH: 1349
; TYPE: PRN
; ORGANISM: Homo sapiens
; US-10-124-805-573

Alignment Scores:
Pred. No.: 1.03e-108 Length: 1349
Score: 1223.00 Matches: 577
Percent Similarity: 59.07% Conservative: 150
Best Local Similarity: 48.36% Mismatches: 380
Query Match: 17.26% Indels: 207
DB: 14 Gaps: 115

US-09-602-362E-22 (1-4115) x US-10-124-805-573 (1-1349)

| | | | |
|----|-----|---|-----|
| QY | 96 | CAAGGGAAGAGACATCAACTTAATATACAGAG---CCAGAGAGAGCTGCTTAAC | 152 |
| DB | 11 | LysargylsThrIleAsnLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu-- | 29 |
| QY | 153 | TGGGCTGGTCAATGGCCT-----GAGGAAGTAGTAACATTTC---GGTA | 194 |
| DB | 30 | -----HisrPalAcysValAsnGlyHisGluGluValValThr-PheLeuValA | 46 |
| QY | 195 | GACAGAGTCCAGCTGAGTCCTTG---TGGCGAACCGGACACCTCTGATGAAGCC-T | 250 |
| DB | 46 | spArgLysCysGlnProAspValLeuAspGlyGluHisArgThrProLeuMetLysAlaL | 66 |
| QY | 251 | TACAAATGCCAGAGGCTTGGC---AAATATTTGATAGATTC-GTGCCGCGATATAATC | 306 |
| DB | 66 | euGlnCysHisGlnGluAlaCysAlaAsnIleLeuIleAspSerGlyAlaAspIleAsnL | 86 |
| QY | 307 | TCGTG---ATGTGATGCAACATCGCTTCCATTATGCGTTTATAGTGGATTGTCAGT- | 362 |
| DB | 86 | euValAspValTyrGlyAsnMetAlaLeuHisTyrAlaValTyrSerGluLeuSerV | 106 |
| QY | 363 | --GGTGGCAAACTGCTCTC-CATGGTCAGTATCGAAGTGC---CAACAAGGCTGCCTCA | 416 |
| DB | 106 | alValAlaLysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeuT | 126 |
| QY | 417 | CACCATTCTTACTACCAT---AAGCAAGAAGTGAACAATTGTGGAATTT---GCTGAT | 470 |
| DB | 126 | hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuLeuIle | 145 |
| QY | 471 | AAATATCAATCGAATGGTTA---ATAAGTTAAATGCACACCTCATGCTGCTGAT | 527 |
| DB | 146 | LysAsnAlaAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAlaVa | 165 |
| QY | 528 | G---TCTGGATCATCAGAGATAGTTGCATGCTTCTTAGCAAAA---TGTGACGCTCTTT-- | 579 |
| DB | 165 | lCysHisGlyLeuSerGluLeuValGlyMetLeuLeuGlnGlnAsnValAspValPheal | 185 |
| QY | 580 | -GTGCAGATTATGAGGAGTAACTCAGAACAT---ATGCTGTACTGTGGATTCTCACA | 635 |
| DB | 185 | aAlaAspIle-CysGlyValThrAlaGluHisTyrAlaValThrCysGlyPheHisIst | 205 |
| QY | 636 | TTCAATAACAAA---TTATGAATATATACGAATATATCAAAATCA---TCAAAATACCAA | 689 |
| DB | 205 | leHisGluGlnIleMet-GlutryIleArgLysLeuSer-LysAsnHisGlnAsnThrAs | 224 |
| QY | 690 | T---CAGAAGAACTCTGCAGGAAACCTGATGAGGTGCACCTC---GGCGGAAGACAC | 743 |
| DB | 224 | nProGluGlyThrSerAlaGlyThrProAspGluAlaAla-ProLeuAlaGluArgThrP | 244 |
| QY | 744 | CTCACGCTGAAGCT---GGTGGAAAAACACCTGATGAGGTGCACCTC---GGTGA | 797 |
| DB | 244 | roAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAla-ProLeuValGlu | 263 |

| | | | |
|----|------|---|------|
| QY | 798 | AAGACACCTGACAGGCTGAAGACT---GGTGGAAAAACACCTGATGAGGCTGCATCT-- | 851 |
| DB | 264 | ArgThrProAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAlaSerLeu | 283 |
| QY | 852 | TGGTGGAGGAACATCTGACAAATTCATGTTGGAGAAAG---ACATCTGGAAGTTCGA | 908 |
| DB | 284 | Val-GluGlyThrSerAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheGl | 303 |
| QY | 909 | ACGTTCAGC---AGAAAAACACCTAGGAATTTACGATCC---TGCAAAAAACACATCTGAG | 962 |
| DB | 303 | uGln-SerAlaGluGluThrProArgGluIleThrSerProAlaLysGluThrSerGluL | 323 |
| QY | 963 | AATTTCAG---GGCCAGCAAGGAAGACCTAGAGATTCGACGGAGAA---AAAGAAGAC | 1016 |
| DB | 323 | ysPheThrTrpPro-AlaLysGlyArgProArgLysIleAlaTrpGluLysLysGluAsp | 342 |
| QY | 1017 | ACACCTAGGGAATATGAGTCCCGAAAAAGAACACCTGAGAAA-----TTA | 1061 |
| DB | 343 | ThrProArgGluIleMetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAla | 362 |
| QY | 1062 | CGTGGCGAGAAAAGGAAGACCTAGGAAGACGCTAGGAGAAAGAAAAAGAAA---CCCTGTAAA | 1118 |
| DB | 363 | LysGlyArgProArgLysIle-----AlaTrpGluLysLysGluThrProValLys | 379 |
| QY | 1119 | GATGGATCGGT-CCAGAGTAAACATCTAATAAATAAAG---TTTTGAAAAAGGAATCT | 1174 |
| DB | 380 | ThrGlyCysValAlaArgValThrSerAsnLysThrLysValLeuGluLysGlyArgSer | 399 |
| QY | 1175 | AAGATGTT-GCATGTCCACAAAAG---AATCATCTACAAAGCAAGTGCCTCATGACAGG | 1230 |
| DB | 400 | LysMetIleAlaCysProThrLysGluSerSerThrLysAlaSerAlaAsnAspGlnArg | 419 |
| QY | 1231 | ---TCCCATCAGATCCAAACAAGGAGATGAAGAAATATCTTT-GATTCTCGGTCCTT | 1286 |
| DB | 420 | PheProSerGluSerLysGlnGluAspGluGluTyrSerCysAspSerArg---Ser | 438 |
| QY | 1287 | TTGT-----AAGTCTCGCAAGATTCAAGTG---GTATACCTCAGTCTATATATAAAGT | 1337 |
| DB | 439 | LeuPheGluSerSerAlaLysIleGlnValCysIleProGluSerIleThrGlnLysVal | 458 |
| QY | 1338 | ANTGAGATAAATAAGAG---TAGAAAGCCTCCTAAAAGCCTCTGCCTTCAAGCGCCA | 1394 |
| DB | 459 | MetGluIleAsnArgGluValGluGluProProLysLysProSerAlaPheLysProAla | 478 |
| QY | 1395 | TTGAAAGCAAAA---CTCTTTCCAAATAGCCTTTG-----AATGAAGAATGAACA | 1442 |
| DB | 479 | IleGluMetGlnAsnSerValProAsnLysAlaPheGluLeuLysAsnGluGln----- | 496 |
| QY | 1443 | AACATTGAAGCAGATCCGTG-TTCCCACCG---AATCCAAACAAGGACTATAAGAAAAT | 1498 |
| DB | 497 | ThrLeuArgAlaAspProMetPheProGluSerLysGlnLysAspTyrGluGluAsn | 516 |
| QY | 1499 | TCTTGGGATTTAGAGTCTCTG---GAGACTGTTTACAGAAG---ATTGTGTTTACCAA | 1552 |
| DB | 517 | SerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLy | 536 |
| QY | 1553 | GGCTACACATCAAAAGAATAAGATAAAA---AATGGAAAA---TAGAAGAGTCCCTAAT | 1606 |
| DB | 536 | saIaThrHisGlnLysGlu-IleAspLysIleAsnGlyLysLeuGluGluSerProAsnL | 556 |
| QY | 1607 | AAAGTGGTCTTCTGAGGCTACCTCGGAAT---GAAATTTCTATTCCACTAAAGCCT--- | 1660 |
| DB | 556 | ysAspGlyLeuLeuLysAlaThrCys-GlyMetLysValSerIleProThrLysAlaLeu | 575 |
| QY | 1661 | AGAATTGAAGACATCAAACTTTTCAAGCGAGCCTCC---GGGAAGCCATCTGCT---TC | 1714 |
| DB | 576 | GluLeuLysAspMetGlnThrPheLysAlaGlu-ProProGlyLysProSerAlaPheGl | 595 |
| QY | 1715 | GAGCCTCCACTGAAATCAAAAGTCTGTCCCAATTAAGCCTTTGG-----AATGAAAA | 1765 |
| DB | 595 | uProAlaThrGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu-- | 614 |

